

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: October 24, 2004, 10:16:46 ; Search time 6393 Seconds

(without alignments)
11325.786 Million cell updates/sec

Title: US-09-977-418-7

Perfect score: 1987

Sequence: 1 tgaatgaagcctctcatg.....cnnncttacttgaagaaa 1987

Scoring table: IDENTITY NUC

Searched: 32822875 seqs, 18219855908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: gb_est1.*
2: gb_est2.*
3: gb_hrc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	819.2	41.2	878	5	EX402471
2	818	41.2	1139	4	BM926609
3	804.4	40.5	901	7	CP552903
4	742.6	37.4	848	7	CP552003
5	643.2	32.4	690	5	BM006459
6	642.2	32.3	974	2	BE250170
7	638.8	32.1	937	5	EX390404
8	624.6	31.4	726	2	BE792513
9	613.8	30.9	681	6	CA424992
10	613.8	30.9	681	5	BM996902
11	613.2	30.9	676	5	BU624626
12	593	29.8	636	2	BE306561
13	587.4	29.6	654	4	BM720799
14	587.4	29.6	976	2	BE203719
15	585	29.4	956	2	BE203383
16	566	28.5	607	5	BM985044
17	560	28.2	594	6	CA424154
18	557.8	28.1	595	1	AT479356
19	538	27.1	562	4	BT113724
20	535	26.9	553	1	AT1079676
21	530.4	26.7	647	5	BO637483
22	527.2	26.5	533	1	AT709377
23	520.8	26.2	937	5	BU178934
24	520.4	26.2	1118	4	BM547264

25	520	26.2	824	7	CP552730
26	519	26.1	530	1	AJ708019
27	511	25.7	554	1	AJ709026
28	503.4	25.3	624	2	BE250457
29	502.6	25.3	605	2	AM844208
30	500.8	25.2	531	1	AJ709551
31	500.4	25.2	528	1	AJ711177
32	492.6	24.8	951	2	BE306631
33	489.8	24.7	518	1	AI1334158
34	483	24.3	533	1	AJ709557
35	480.4	24.2	508	6	CB267862
36	478.4	24.1	495	1	AI041115
37	475.8	23.9	506	1	AJ707934
38	473.4	23.9	478	1	AM688131
39	475.2	23.9	481	1	AI104359
40	474	23.9	528	1	AJ711705
41	473.4	23.8	505	2	AM082736
42	473.2	23.8	532	1	AJ709542
43	471	23.7	488	1	AI133328
44	470	23.7	833	7	CP552429
45	465.2	23.4	518	6	CD625679

ALIGNMENTS

RESULT 1
BX402471/c
LOCUS
DEFINITION
BX402471 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSOD1039B13 3-PRIME, mRNA sequence.
ACCESSION
BX402471
VERSION
BX402471.2 GI:46874966
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
Li W.B., Gruber C., Jessee J. and Polayes D.
Full-length cDNA libraries and normalization
1 (bases 1 to 878)
Unpublished (2001)
JOURNAL
On May 13, 2003 this sequence version replaced gi:30618825.
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
545_r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?c=CS1A1010F0FNPLKc=545_r.
Location/Qualifiers

FEATURES

source
1..878
/organism="Homo sapiens"
/mol_type="RNA"
/db_xref="taxon:9606"
/clone="CSOD1039B13"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 41.2%; Score 819.2; DB 5; Length 878;
Best Local Similarity 96.9%; Pred. No. 6e-211;
Matches 854; Conservative 1; Mismatches 22; Indels 4; Gaps 2;

QY 1013 GGGTTTTTCCATACAGCTGTTGTAAGAAAAGTGCAGTAACTCTGCGCTCTGCGCC 1072
 Db 878 GGGTTTTTCCATACAGCTGTTGTAAGAAAAGTGCAGTAACTCTGCGCTCTGCGCC 819
 QY 1073 TGGTGGGGCTTCCAGGAAAGGCAAGTGGGGTTGGGGGCTGGGCTTCTTCTCTCC 1132
 Db 818 TGGTGGGGCTTCCAGGAAAGGCAAGTGGGGTTGGGGGCTGGGCTTCTTCTCTCC 759
 QY 1133 CACAGGCTGTGTTCTTCTGTTGGGGGCTGCTCCCATGAGACAGAGTACCTTAACAGATGGA 1192
 Db 758 CACAGGCTGTGTTCTTCTGTTGGGGGCTGCTCCCATGAGACAGAGTACCTTAACAGATGGA 699
 QY 1193 GCCAGGCGATGATGAGGGGCTTTGGGCTCTGAGAGTTGAACCCAGCTTCTTCCACTTC 1252
 Db 698 GCCAGGCGATGATGAGGGGCTTTGGGCTCTGAGAGTTGAACCCAGCTTCTTCCACTTC 639
 QY 1253 CCTCCGGGAGTGAAGCTTCCATCCATCCCTCTTATCTATGATATATAGGCTGG 1312
 Db 638 CCTCCGGGAGTGAAGCTTCCATCCATCCCTCTTATCTATGATATATAGGCTGG 579
 QY 1313 TGTGTATACACACACACCCCTATCGTTGTCCTTCAATATCTGAGCATTCATGATG 1372
 Db 578 TGTGTATACACACACACCCCTATCGTTGTCCTTCAATATCTGAGCATTCATGATG 519
 QY 1373 GGGCAATTCAGAGCTTTCTCAATACAGTTCATTCATTCATTCATTCATTCATTCATTC 1432
 Db 518 GGGCAATTCAGAGCTTTCTCAATACAGTTCATTCATTCATTCATTCATTCATTCATTC 459
 QY 1433 CATCCCGAGCAGTGTGCTGTTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1492
 Db 458 CATCCCGAGCAGTGTGCTGTTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 399
 QY 1493 TCAACGCTGCTTCAACTGCCCACTCTGGGCACTGAGAGTATTTCCCTCATTTTAC 1552
 Db 398 TCAACGCTGCTTCAACTGCCCACTCTGGGCACTGAGAGTATTTCCCTCATTTTAC 339
 QY 1553 CTCTTAAGGCTATGACCCCTCCAGGCTTCCAGCTGAGGAGTGGGGGAGTATAG 1612
 Db 338 CTCTTAAGGCTATGACCCCTCCAGGCTTCCAGCTGAGGAGTGGGGGAGTATAG 279
 QY 1613 GAAAGCCCCCATCTCCATCTGGGATAGGAGCTTCCATCAGCTTAACTCTGGGAAAT 1672
 Db 278 GAAAGCCCCCATCTCCATCTGGGATAGGAGCTTCCATCAGCTTAACTCTGGGAAAT 219
 QY 1673 GCGTGTGCCCCAGGAGCTTGGTTGGTCTCCCATACAGAGGAGGAGTGGAGGG 1732
 Db 218 GCGTGTGCCCCAGGAGCTTGGTTGGTCTCCCATACAGAGGAGGAGTGGAGGG 159
 QY 1733 AAGGATGAGTCTCAGTTAGCAGGGGATCCCAAGGAGTCAAGCTCTCTCCATGCT 1792
 Db 158 AAGGATGAGTCTCAGTTAGCAGGGGATCCCAAGGAGTCAAGCTCTCTCCATGCT 99
 QY 1793 CTCTGTGAGTGTGCTTGGGAGTGGTCTCTCACTCCCACTCTGGG-CCCTTGGGGA 1851
 Db 98 CTCTGTGAGTGTGCTTGGGAGTGGTCTCTCACTCCCACTCTGGG-CCCTTGGGGA 39
 QY 1852 GGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1892
 Db 38 GGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1

RESULT 2
 LOCUS BM926609 1139 bp mRNA linear EST 12-MAR-2002
 DEFINITION AGENCOURT 6644689 NIH_MGC_122 Homo sapiens cdna IMAGE:5767017
 5' UTR sequence.
 ACCESSION BM926609
 VERSION BM926609.1 GI:19376988
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 AUTHORS NIH-MGC
 TITLE NIH-MGC
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 DNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LLM12824 row: 3 column: 10
 High quality sequence stop: 717.
 Location/Qualifiers

FEATURES
 source
 1. 1139
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5767017"
 /lab_host="DH10B"
 /clone_1lb="NIH MGC 122"
 /note="Organ: pooled lung and spleen; Vector: pCMV-Sport6;
 Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
 anonymous pool of 24 week female lung, 16 week female
 spleen, and 20-22 week male spleens. Library is oligo-dT
 primed and directionally cloned (EcoRV site is destroyed
 upon cloning). Average insert size 1.4 kb, insert size
 range 1-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 026. Note:
 this is a NIH_MGC Library."

ORIGIN

Query Match 41.2%; Score 818; DB 4; Length 1139;
 Best Local Similarity 93.3%; Pred. No. 1,4e-210;
 Matches 941; Conservative 0; Mismatches 55; Indels 13; Gaps 8;

QY 605 ACTTAATCCAGATGACGGGGCTGGGGTCAAGGAGAAAGATTAAAGAACCCACTGTGG 664
 Db 12 ACTTAATCCAGATGACGGGGCTGGGGTCAAGGAGAAAGATTAAAGAACCCACTGTGG 71
 QY 665 GTCAGGGGAATGGGACAGAGAGACATATGGGCAAGCTCTGACAGACAGACAGACAGAC 724
 Db 72 GTCAGGGGAATGGGACAGAGAGACATATGGGCAAGCTCTGACAGACAGACAGACAGAC 131
 QY 725 AACCTCTGATCTATGAAGTCTCTGAGGGCAAGGGACCAAGGACCTGGAACCTCTTG 784
 Db 132 AACCTCTGATCTATGAAGTCTCTGAGGGCAAGGGACCAAGGACCTGGAACCTCTTG 191
 QY 785 GCCAAGGGGAGTGGGAGAGCAGAGGGAAGTCAAGGGCAAGGGTGCCTATCTAAGTGA 844
 Db 192 GCCAAGGGGAGTGGGAGAGCAGAGGGAAGTCAAGGGCAAGGGTGCCTATCTAAGTGA 251
 QY 845 ACTAATGGCCGAGGGCTCAGCAAGGCCCAAGAGAGACAGCCGTGACGGTAAACCTTCCC 904
 Db 252 ACTAATGGCCGAGGGCTCAGCAAGGCCCAAGAGAGACAGCCGTGACGGTAAACCTTCCC 311
 QY 905 TCTAACAGGCTCCAAAGCCCTCAGCCAGGAGAGGCTGCTGCCACCCCGTCCCGCAG 964
 Db 312 TCTAACAGGCTCCAAAGCCCTCAGCCAGGAGAGGCTGCTGCCACCCCGTCCCGCAG 371
 QY 965 CCAAGCTGCTGCTCCAGGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1024
 Db 372 CCAAGCTGCTGCTCCAGGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 431
 QY 1025 TACAGCTGCTGTGTAAGAACTGATGAATCTCTGCGCTCTGCGCTCTGCGGACCTTC 1084
 Db 432 TACAGCTGCTGTGTAAGAACTGATGAATCTCTGCGCTCTGCGCTCTGCGGACCTTC 491
 QY 1085 CAGGCAAGGCAAGTGGGGGTTGGGGGTTGGGGGTTGGGGGTTGGGGGTTGGGGGTTGGGG 1144

Db 492 CAGGCAAGGCCACGATGGGGGTTGGGGGCTGCTCTTCCCTCCACAGGCTGTG 551
 QY 1145 TTGNTGGGGGTGCTCCCATGACAGACAGATCACCTAACAGAGATGAGCCAGGCGATG 1204
 Db 552 TTTCTGGGGCTGCTCCCATGACAGACAGATCACCTAACAGAGATGAGCCAGGCGATG 611
 QY 1205 ATGGGGCTTGGGTCTCGAGGTGGAACCCAGCTTCTTGCCACCTTCCCTCC -GGCAG 1263
 Db 612 ATGGGGCTTGGGTCTCGAGGTGGAACCCAGCTTCTTGCCACCTTCCCTCC -GGCAG 671
 QY 1264 TCAGCTCTCATCATCTCCCTCTTTAATGATCTAAGGCTGGT -GGTGTAC 1322
 Db 672 TCAGCTCTCATCATCTCCCTCTTTAATGATCTAAGGCTGGTGGGTGTAC 731
 QY 1323 A-ACACACCCCTATCGTGTCTTCAAAATACAGATTA-CCATTGTTGAGCCCAAT 1380
 Db 732 ACACACACCCCTATCGTGTCTTCAAAATACAGATTA-CCATTGTTGAGCCCAAT 791
 QY 1381 TCAGAGCTTTCTCAATCAGATTTTCAATCTCCATTTCTTAAAGGGAATCTCCCG 1440
 Db 792 TCAGAGCTTTCTCAATCAGATTTTCAATCTCCATTTCTTAAAGGGAATCTCCCG 851
 QY 1441 AGCAGCTGAGTGTCTGCTTGTCTAC--TGAAGTTAGATCTGAA---CCCAGGCTGCA 1495
 Db 852 AGCAGCTGAGTGTCTGCTTGTCTAC--TGAAGTTAGATCTGAA---CCCAGGCTGCA 911
 QY 1496 ACNGCTGCTCTCACTCCCACTCTGGGCA--CTGAGAGTATTTTCCCTCATTC--TA 1551
 Db 912 CAGCTGCTCTCACTCCCACTCTGGGCACTGGAAGAGATTTTCCCTCATCTCA 971
 QY 1552 CCTCTCAAGCTATGACACCCCTCCCAAGCTCTTCAAGCTGGGGATG 1600
 Db 972 CCTCTCAAGCTATGACACCCCTCCCAAGCTCTTCAAGTGTGGGG 1020

RESULT 3
 CF552903 901 bp mRNA linear EST 22-SEP-2003
 LOCUS ABENECOURT 15596050 NIH_MGC_183 Homo sapiens CDNA clone
 DEFINITION IMAGE:30529648 5', mRNA sequence.
 ACCESSION CF552903
 VERSION CF552903.1 GI:34889737
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 901)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: gcaps-remail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein
 CDNA Library Preparation: Invitrogen Corp
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: NDAM617 row: 1 column: 17
 High quality sequence stop: 674.
 Location/Qualifiers

FEATURES
 Source
 1..901
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30529648"
 /lab_host="DH10B-Tona (T1 and T5 phage resistant)"
 /clone_lib="NIH_MGC_183"

/note="Organ: Pooled muscle (cardiac and skeletal);
 Vector: pCMV-Sport6.1; Site 1: EcoRV (destroyed); Site 2:
 NotI; Library is oligo-AT primed and directionally cloned
 (EcoRV site is destroyed upon cloning). Average insert
 size 1.7. Library was constructed by Invitrogen."

Query Match 40.5%; Score 804.4; DB 7; Length 901;
 Best Local Similarity 97.1%; Pred. No. 6.3e-207;
 Matches 872; Conservative 0; Mismatches 18; Indels 8; Gaps 5;

QY 760 GACACAGGAGCTGGAACCTCTTGCCCAAGGGAGTGGGAGAGCAGAGGAGGTGAC 819
 Db 1 GACACAGGAGCTGGAACCTCTTGCCCAAGGGAGTGGGAGAGCAGAGGAGGTGAC 59
 QY 820 AGGCAAGGGGTGCTATCTAAGTGAATTAATGCTCCAGAGGCTCAGCAAGGCCAAGAGA 879
 Db 60 AGGCAAGGGGTGCTATCTAAGTGAATTAATGCTCCAGAGGCTCAGCAAGGCCAAGAGA 119
 QY 880 GACAGCGTGAAGCTGAATCTTCCCTCTACAGAGCTCCAAAGCCCAAGCCAGAGAGAG 939
 Db 120 GACAGCGTGAAGCTGAATCTTCCCTCTACAGAGCTCCAAAGCCCAAGCCAGAGAGAG 179
 QY 940 CTGCTGACCCAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 999
 Db 180 CTGCTGACCCAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 239
 QY 1000 CTGTAATATGATGGGGTTTTTTCATATAGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1059
 Db 240 CTGTAATATGATGGGGTTTTTTCATATAGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 299
 QY 1060 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1115
 Db 300 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 359
 QY 1120 TCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1179
 Db 360 TCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 419
 QY 1180 AACAGAGTGAAGCCAGGCGATGATGGGCTTGGGCTCTCGAGGTTGAGACCCCACT 1239
 Db 420 AACAGAGTGAAGCCAGGCGATGATGGGCTTGGGCTCTCGAGGTTGAGACCCCACT 479
 QY 1240 TCTTCCAGCTTCCCTCC -GGCAGTCACTCTCCATCCATCCCTCTTTAATCTATGA 1298
 Db 480 TCTTCCAGCTTCCCTCC -GGCAGTCACTCTCCATCCATCCCTCTTTAATCTATGA 539
 QY 1299 ACTTATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1357
 Db 540 ACTTATAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 599
 QY 1358 ATTACCATTTGTTGAGGCGCAATTGAGGCTTTCGAATCAGATTCAATCTCATTT 1417
 Db 600 ATTACCATTTGTTGAGGCGCAATTGAGGCTTTCGAATCAGATTCAATCTCATTT 659
 QY 1418 TCATTAAAGGGGAAACATCCCGAGCCACTGAGTCTGTGCTTGTCTGCTGCTGCTGCTGCT 1477
 Db 660 TCATTAAAGGGGAAACATCCCGAGCCACTGAGTCTGTGCTTGTCTGCTGCTGCTGCTGCT 719
 QY 1478 TCTGAACCCAGGCTGCTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1537
 Db 720 TCTGAACCCAGGCTGCTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779
 QY 1538 TTCCCTCATTTCTACCTCTTAAGGCTATGACCCCTCCCAAGCTTTCAGAGTGGG - 1595
 Db 780 TTCCCTCATTTCTACCTCTTAAGGCTATGACCCCTCCCAAGCTTTCAGAGTGGG 839
 QY 1596 --GATGGGGAGTCTATAGAGAAAGCCCTC-ATCTCCATCTGGGATAGGACCTTC 1650
 Db 840 AAGGGGGGGGATCATAGAGAAAGCCCTC-ATCTCCATCTGGGATAGGACCTTC 897

RESULT 4

CF552003 848 bp mRNA linear EST 22-SEP-2003
 LOCUS AGENCOURT 15595803 NIH MGC 183 Homo sapiens cDNA clone
 DEFINITION IMAGE:30529976 5', mRNA sequence.
 ACCESSION CF552003
 VERSION CF552003.1 GI:34888837
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 848)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 AUTHORS Unpublished (1999)
 JOURNAL Contact: Daniela S. Gerhard, Ph.D.
 COMMENT Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein
 cDNA Library Preparation: Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: NDAM618 row: 9 column: 09
 High quality sequence start: 2
 High quality sequence stop: 666.
 Location/Qualifiers
 1..848
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30529976"
 /lab_host="DH10B-Tona (T1 and T5 phage resistant)"
 /note="Organ: Pooled muscle (cardiac and skeletal);
 Vector: PCMV-SPORT6.1; Site 1: EcoRV (destroyed); Site 2:
 NotI; Library is oligo-dT primed and directionally cloned
 (EcoRV site is destroyed upon cloning). Average insert
 size 1.7. Library was constructed by Invitrogen."
 ORIGIN
 Query Match 37.4%; Score 742.6; DB 7; Length 848;
 Best Local Similarity 96.2%; Pred. No. 3.9e-190;
 Matches 816; Conservative 0; Mismatches 20; Indels 12; Gaps 5;
 QY 516 CCAACTAGGCGCTTCACCTATCCAGGCGAGGCTCCCTGACACATG 575
 DB 1 CCCACTAGGCGCTTCACCTATCCAGGCGAGGCTCCCTGACACATG 60
 QY 576 ATGCCAGATTCAGATTGGCTCGCTACTTATCCAGAGTACAGGGGCTGGGCTAG 635
 DB 61 ATGCCAGATTCAGATTGGCTCGCTACTTATCCAGAGTACAGGGGCTGGGCTAG 120
 QY 636 GGAAGGAGATCTAAGAACCCACTGTGGGTGAGGGAATGGGACACAGACATATG 695
 DB 121 GGAAGGAGATCTAAGAACCCACTGTGGGTGAGGGAATGGGACACAGACATATG 180
 QY 696 GGAAGCTTCGACGACAGACAGACAAACCTCTGATCTATGAATCTCTGACGGGC 755
 DB 181 GGAAGCTTCGACGACAGACAGACAAACCTCTGATCTATGAATCTCTGACGGGC 240
 QY 756 AAGGGAGACGAGGAGCTGGAACCTCTTGGCCAAAGGGAGTGGAGAGAGGGAAG 815
 DB 241 AAGGGAGACGAGGAGCTGGAACCTCTTGGCCAAAGGGAGTGGAGAGAGGGAAG 300
 QY 816 TACAGGCAAGGGTGCCTATCTAATGTGAATTAATGCTCCAGAGGCTCAGACGCAAG 875
 DB 301 TACAGGCAAGGGTGCCTATCTAATGTGAATTAATGCTCCAGAGGCTCAGACGCAAG 360

QY 876 AGGAGACAGCGGTGACGGTAACTTCCTCTACACAGCTCCAAAGCCCAAGCCAGCGAG 935
 DB 361 AGGAGACAGCGGTGACGGTAACTTCCTCTACACAGCTCCAAAGCCCAAGCCAGCGAG 420
 QY 936 CAGGCTGCGTCCCGACCGCGTCCCGACGACCTGCTGCGACGAGACCATATGCC 995
 DB 421 CAGGCTGCGTCCCGACCGCGTCCCGACGACCTGCTGCGACGAGACCATATGCC 480
 QY 996 ACATCTATATATAGATGGGTTTTTCCATATACAGTGTTCGTGAATAAATGATGAAC 1055
 DB 481 ACATCTATATATAGATGGGTTTTTCCATATACAGTGTTCGTGAATAAATGATGAAC 540
 QY 1056 TCCGCGCGTCCCGACCGCGTCCCGACGACCTGCTGCGACGAGACCATATGCC 1114
 DB 541 TCCGCGCGTCCCGACCGCGTCCCGACGACCTGCTGCGACGAGACCATATGCC 600
 QY 1115 GCTGTCCTTCCTCCCTCCACAGAGCTGTGTTGTTGAGGCTGCTCCATGACAGAGAT 1174
 DB 601 GCTGTCCTTCCTCCCTCCACAGAGCTGTGTTGTTGAGGCTGCTCCATGACAGAGAT 660
 QY 1175 CACCTAACAGAGATGGAAGCCAGGCGATGATGGGG--CTTGGTCTTCGAGTTGAC 1232
 DB 661 CACCTAACAGAGATGGAAGCCAGGCGATGATGGGGCTTTGGGGTCTCGAGTTGAG 720
 QY 1233 CCAGCTTCTTGCACCTTCCC--TCCGCACTGAGCTCTCCATCCATCCCTCTTTA 1290
 DB 721 CCAGCTTCTTGCACCTTCCCCTCCGAGGAGTCACTCTCATCATCCCTCTTTA 780
 QY 1291 ATCTATGATCTATAGGCTC-----GATGTGTAAACAACACCCCTATC--GTTGCC 1343
 DB 781 ATCTATGATCTATAGGCTCCTCGGGGATGTGTAAACAACCCCTATCCTGTTGCC 840
 QY 1344 TTCAATA 1351
 DB 841 TTCAATA 848
 RESULT 5
 BQ006459/c 690 bp mRNA linear EST 26-MAR-2002
 LOCUS UR-H-B11-424-K-12-0-UI.81 NCI CGAP_B11 Homo sapiens cDNA clone
 DEFINITION IMAGE:5846147 3', mRNA sequence.
 ACCESSION BQ006459
 VERSION BQ006459.1 GI:19731359
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 690)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgabs-remail.nih.gov
 Tissue Procurement: Dr. Jose Mercuende
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
 Seq primer: M13 FORWARD
 POLVA=Yes.
 Location/Qualifiers
 1..690
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5846147"
 /tissue_type="Chondrosarcoma"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"

/clone_lib="NCI_CGAP_E11"
 /note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)
 with a modified polylinker; Site 1: EcoR I; Site 2: Not I;
 NCI CGAP E11 is a normalized cDNA library containing the
 following tissue(s): Chondrosarcoma. The library was
 constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into pT73-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 ACACTTGAC.
 TAG_TISSUE=chondrosarcoma
 TAG_LIB=UI-H-E11
 TAG_SEQ=ACACTTGAC"

ORIGIN

Query Match 32.4%; Score 643.2; DB 5; Length 690;
 Best Local Similarity 99.0%; Pred. No. 3,8e-163;
 Matches 667; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

1252 CCCCTCCGCGAGTCAGCTCCATCCATCCCTCTTATATATATATATATAGAGCTCG 1311
 689 CCTCCGCGAGTCAGCTCCATCCATCCCTCTTATATATATATATATAGAGCTCG 630

1312 GTGTGTGTATACA-ACAACCCCTATGCTTGTCCCTTAATTAATCAATCAATGATTGTT 1370
 629 GTGTGTGTATACAACAACCCCTATGCTTGTCCCTTAATTAATCAATCAATGATTGTT 570

1371 GAGGCCAAATTCAGAGTTTCTCAATCAATTAATTAATCAATCAATCAATCAATCAATCAAT 1430
 569 GAGGCCAAATTCAGAGTTTCTCAATCAATTAATTAATCAATCAATCAATCAATCAATCAAT 510

1431 AATATCCCGAGCAGCTAGTGTGCTTGTCTCACTGAAGTTAATCAATCAATCAATCAATCAAT 1490
 509 AATATCCCGAGCAGCTAGTGTGCTTGTCTCACTGAAGTTAATCAATCAATCAATCAATCAAT 450

1491 TGTCAACNGCTGCTCTCAATCTCCCACTCTGAGCACTGAAGTTAATCAATCAATCAATCAAT 1550
 449 TGTCAACNGCTGCTCTCAATCTCCCACTCTGAGCACTGAAGTTAATCAATCAATCAATCAAT 390

1551 ACCCTCTAAGGCTATGACACCCCTCCCACTCTGAGCTGAGGAGTGGAGGAGTCAAT 1610
 389 ACCCTCTAAGGCTATGACACCCCTCCCACTCTGAGCTGAGGAGTGGAGGAGTCAAT 330

1611 AGGAAAAGCCCCCATCTCCCATCTGGATAGGAGACCTTCATCAAGCTTAACTTAACTTAACTTAA 1670
 329 AGGAAAAGCCCCCATCTCCCATCTGGATAGGAGACCTTCATCAAGCTTAACTTAACTTAACTTAA 270

1671 ATGCTCTGCTGCCCCCAATGACTCTTGTGTTCTGTCTCCCATCAAGAGAGAGGAGTGGAGG 1730
 269 ATGCTCTGCTGCCCCCAATGACTCTTGTGTTCTGTCTCCCATCAAGAGAGAGGAGTGGAGG 210

1731 GGAAGGAGTGGAGTCTCAATTAAGCAGAGGAGTCCCAAGGAGAGTCAAGCTCTCCCTCAATGAG 1790
 209 GGAAGGAGTGGAGTCTCAATTAAGCAGAGGAGTCCCAAGGAGAGTCAAGCTCTCCCTCAATGAG 150

1791 CTCTCTGTGATCAATGTGCTTAAAGGTGAGCTCTCACTCCCACTCACTGGG-CCCTTGGGG 1849
 149 CTCTCTGTGATCAATGTGCTTAAAGGTGAGCTCTCACTCCCACTCACTGGG-CCCTTGGGG 90

1850 GAGAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1909
 89 GAGAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 30

1910 GACAGTCTCAAGA 1923
 29 GACAGTCTCAAGA 16

RESULT 6
 BE250170 974 bp mRNA linear EST 13-JUN-2000
 LOCUS 600943555P1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960076 5',
 DEFINITION mRNA sequence.
 ACCESION BE250170.1 GI:9120277
 VERSION BE250170
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NIH-MGC http://mgc.nci.nih.gov/
 1 (bases 1 to 974)
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
 Plate: L16M52 row: n column: 13
 High quality sequence stop: 692.
 Location/Qualifiers
 1..974
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2960076"
 /tissue_type="rhabdomyosarcoma"
 /lab_host="MDH10B (phage-resistant)"
 /clone_lib="NIH_MGC_17"
 /note="Organ: muscle; Vector: pOT57; Site 1: EcoRI;
 Site 2: XhoI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoR/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

FEATURES

ORIGIN

Query Match 32.3%; Score 642.2; DB 2; Length 974;
 Best Local Similarity 96.0%; Pred. No. 7.7e-163;
 Matches 700; Conservative 0; Mismatches 25; Indels 4; Gaps 4;

782 TTGGCCAAAGGAGTGGAGAGACAGAGGAAAGTCAAGGCAAGGAGTGGCTTAACT 841
 2 TTGGCCAAAGGAGTGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTAAAGT 61

842 GGAAGTAAATTTGGCCGAGGAGTCAAGCAAGGCCAAGAGAGACAGCCGTGACGGTAAACTTC 901
 62 GGAAGTAAATTTGGCCGAGGAGTCAAGCAAGGCCAAGAGAGACAGCCGTGACGGTAAACTTC 121

902 CCTCTACAGAGCTCAAGGCCCAAGGCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 961
 122 CCTCTACAGAGCTCAAGGCCCAAGGCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 181

962 CAGCCAGTGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1021
 182 CAGCCAGTGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 241

1022 CAATACAGTGGTGGTGAATAAGTCAATGAATCTGCGCTGCTGCGCTGCTGCGGAG 1081
 242 CAATACAGTGGTGGTGAATAAGTCAATGAATCTGCGCTGCTGCGCTGCTGCGGAG 301

1082 CTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1141
 302 CTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 361

QY 1142 GTGTTCTGGGGCTGCTCCCATGACAGACAGATCACTTAACAGAGATGAGAGGCA 1201
 Db 362 GTGTTCTGGGGCTGCTCCCATGACAGACAGATCACTTAACAGAGATGAGAGCA 420
 QY 1202 TGGATGGGGCTTGGGCTCTGAGGTTGAGACCCCACTTCTGCACTTCCCTCC 1260
 Db 421 TGGATGGGGCTTGGGCTCTGAGGTTGAGACCCCACTTCTGCACTTCCCTCC 480
 QY 1261 CAGTCAGCTCTTCATCATCATCCCTCTTAACTAATGAATCTAAGGCTGGTGTGA 1320
 Db 481 CAGTCAGCTCTTCATCATCATCCCTCTTAACTAATGAATCTAAGGCTGGTGTGA 540
 QY 1321 ACA-AACACACCCCTATGCTTGTCTTCAAACTACTGACATTAACATGTTAGGCCAA 1379
 Db 541 ACAACACACCCCTATGCTTGTCTTCAAACTACTGACATTAACATGTTAGGCCAA 600
 QY 1380 TTC-AGAGCTTCTCAATCAATGATTTACATCTCCATTTCTTAACGGGAAATATCC 1438
 Db 601 TTCAGAGCTTCTCAATCAATGATTTACATCTCCATTTCTTAACGGGAAATATCC 660
 QY 1439 CGAGCCACTGAGTGTGCTTGTGCTGACGTAAGGTTAGATCTGAACCCAGGCTGCAACN 1498
 Db 661 CGAGCCACTGAGTGTGCTTGTGCTGACGTAAGGTTAGATCTGACCCGGGTGTCACTGCT 720
 QY 1499 GCTGCTCTC 1507
 Db 721 GCTGCTCTC 729
 RESULT 7
 LOCUS BX390404 937 bp mRNA linear EST 28-APR-2004
 DEFINITION BX390404 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA
 clone CS0D1039YB13 5-PRIME, mRNA sequence.
 ACCESSION BX390404
 VERSION BX390404.2 GI:46846080
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 937)
 Li, W.-B., Gruber, C., Jesse, J., and Polayes, D.
 Full-length cDNA libraries and normalization.
 JOURNAL Unpublished (2001)
 COMMENT On May 8, 2003 this sequence version replaced gi:30461470.
 CONTACT: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 543.r
 For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?e=CS0BA030ZE10_CS02861_1ac=545.r.
 FEATURES
 source
 1..937
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1039YB13"
 /tissue_type="PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo (dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized."
 ORIGIN
 Query Match 32.1%; Score 638.8; DB 5; Length 937;

Best Local Similarity 88.6%; Pred. No. 6,4e-162; Matches 845; Conservative 0; Mismatches 82; Indels 27; Gaps 13;
 QY 9 GGGCTGTCATGCTCTGTCAAG-AGCCATCCCAACCAAGCCCGGT-GGGAGGCAAC 66
 Db 1 GGGCTGTCATGCTCTGTCAAGAGCCATCCCAACCAAGCCCGGTGGGAGGCAAC 60
 QY 67 ATCAAGACCTTAGAGAGCGCTTGAAGTTGGCGGTGAGAGTGAAGCTTCACTGAA 126
 Db 61 ATCAAGACCTTAGAGAGCGCTTGAAGTTGGCGGTGAGAGTGAAGCTTCACTGAA 120
 QY 127 GACATCAATTGTACACACCTTCCAAACCAACCAACGAGGTGGGGGCTGAGAACTGG 186
 Db 121 GACATCAATTGTACACACCTTCCAAACCAACCAACGAGGTGGGGGCTGAGAACTGG 175
 QY 187 CGGCTGAACGACACTGTATGAAACCTTACAGTACAGAGTGGCGGAGTGGCGGA 246
 Db 176 CGGCTGAACGACACTGTATGAAACCTTACAGTACAGAGTGGCGGAGTGGCGGA 227
 QY 247 GGAAGTGAACCCGAGCGGTGAGCTTGTGCTGCTGAGAGAGAGAGAGAGAGAGAG 306
 Db 228 GGAAGTGAACCC--GAGTGGGTGAGCTG--CTGCGGAGAGAGAGAGAGAGAGAG 281
 QY 307 TAGTCCGGGACAGCGGTCAACCCGATACAGAAACGTTCCAGACACTTCCGAGAGAT 366
 Db 282 TAGTCCGGGACAGCGGTCAACCCGATACAGAAACGTTCCAGACACTTCCGAGAGAT 341
 QY 367 CAAATCTGAGTCT 426
 Db 342 CAAATCTGAGTCT 401
 QY 427 AGCTTGTCTTACCCCATTTACAAAGCTCCAGACATCTGAGCCAGGTTTACGCCCAAC 486
 Db 402 AGCTTGTCTTACCCCATTTACAAAGCTCCAGACATCTGAGCCAGGTTTACGCCCAAC 461
 QY 487 GCAACCCAGACCCCAAGTGAACCATCTCCCAACCTAGAGGCGCTCACTCTTACAGAG 546
 Db 462 GCAACCCAGACCCCAAGTGAACCATCTCCCAACCTAGAGGCGCTCACTCTTACAGAG 521
 QY 547 AGCCACAGAGCT 606
 Db 522 AGCCACAGAGCT 581
 QY 607 TTATCCAGAGTACAGGCGTGGGCTGAGGAGAGAGATCTTAAAGACCCACTGTGGGT 666
 Db 582 TTATCCAGAGTACAGGCGTGGGCTGAGGAGAGAGATCTTAAAGACCCACTGTGGGT 641
 QY 667 CAGGGAAATGGGACAGACAGACATATGGGCAAGCTCTGACAGACAGACAGACAA 726
 Db 642 CAGGGAAATGGGACAGACAGACATATGGGCAAGCTCTGACAGACAGACAGACAA 701
 QY 727 CCTCTGATCTATGAAGTCTCTGACAGGCAAGGGGACCA--GGACCTTGAACCTCTTGG 784
 Db 702 CCTCTGATCTATGAAGTCTCTGACAGGCAAGGGGACCAAGGACCTTGAACCTCTTGG 761
 QY 785 GCCAAGGGAG--TGGAGAGACAGAGAGAGGT--CACAGGCAAGGG--TGCTATCTAAGT 841
 Db 762 GCCAAGGGAGTGGAGAGAGAGAGAGAGGTTCACAGGCAAGGGTTCCTATCTTAAG 821
 QY 842 GGAATCTAA-TTGGCCGAGGGCTGACAGAGCCCAAGAGAGACACCTGACCGTAACCTT 900
 Db 822 TGGACTAATTTGGCCGAGGGCTGACAGAGCCCAAGAGAGACACCGGAGCGGAACTT 881
 QY 901 CCCCTTACAGGCTTCAAGGCCCAAGGAGAGAGAGGCTGCTGACCAACCC 954
 Db 882 CCCCTTACAGGCTTCAAGGCCCAAGGAGAGAGAGGCTGCTGACCAACCC 935
 RESULT 8
 LOCUS BF792513 726 bp mRNA linear EST 12-JAN-2001
 DEFINITION BF792513 602253622F1 NIH_MGC_84 Homo sapiens CDNA clone IMAGE:4345876 5',
 mRNA sequence.

ACCESSION BE792513
 VERSION BE792513.1 GI:12097567
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 726)
 REFERENCE NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov
 plate: LIM966 row: 1 column: 05
 High quality sequence stop: 724.
 Location/Qualifiers

FEATURES

Source
 1..726
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4345876"
 /tissue_type="adrenal cortex carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NIH_MGC_84"
 /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site: 1; NotI; Site 2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 1.229 kb. library enriched for full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC library."

ORIGIN

Query Match 31.4%; Score 624.6; DB 2; Length 726;
 Best Local Similarity 97.6%; Pred. No. 4,36-158;
 Matches 686; Conservative 0; Mismatches 11; Indels 6; Gaps 5;
 QY 867 AAGGCCAAGAGGAGACCGCGTGAAGTAACTTCCCTTACCAAGCTTCCAGCCCCAC 926
 DB 27 AAGGCCAAGAGGAGACCGCGTGAAGTAACTTCCCTTACCAAGCTTCCAGCCCCAC 86
 QY 927 GCCAGCAGCAGCGCTGCTGCCACCCCGTGCACCCAGCAGCTGCTGTGCCAGGCGAG 986
 DB 87 GCCAGCAGCAGCGCTGCTGCCACCCCGTGCACCCAGCAGCTGCTGTGCCAGGCGAG 146
 QY 987 AGCCATGCCACATCTGTATATATAGTGGGATTTCATACAGCTGCTGTGAAAACT 1046
 DB 147 AGCCATGCCACATCTGTATATATAGTGGGATTTCATACAGCTGCTGTGAAAACT 206
 QY 1047 GCATGAACCTCTGCTCCGCTGCGGCTGCGGGGCTCCAGGCAAGGCAAGGGGGTTG 1106
 DB 207 GCATGAACCTCTGCTCCGCTGCGGCTGCGGGGCTCCAGGCAAGGCAAGGGGGTTG 266
 QY 1107 GGGGTGGGGCTGTCTCTTCCCTCCACAGGCTGTGTCTGTGGGGCTGTCCCATGCA 1166
 DB 267 GGTGTGGGGCTGTCTCTTCCCTCCACAGGCTGTGTCTGTGGGGCTGTCCCATGCA 326
 QY 1167 GACGAGATCACTTAACAGAGTGAAGCGAGGAGATGAGGGGCTTGGGTCTCTGAGG 1226
 DB 327 GACGAGATCACTTAACAGAGTGAAGCGAGGAGATGAGGGGCTTGGGTCTCTGAGG 385
 QY 1227 TTGAGACCCCACTTCTGCGACCTTCCCTCC-GGCAGTCAAGCTTCCATCCATCCCTT 1285
 DB 386 TTGAGACCCCACTTCTGCGACCTTCCCTCCGGGAGTGAAGCTCTCATCCATCCCTT 445
 QY 1286 CTTTATCTATGAATCTATAGGCTCGGTGTGTGTA--ACAACACACCCCTATGTTTCTC 1343
 DB 446 CTTTATCTATGAATCTATAGGCTCGGTGTGTGTAACACACACCCCTATGTTTCTC 505

QY 1344 TTCAATACTCAGCATTTACCATTTGGTTGAGGCCAAATTACAGAGCTTCTCAATCAGATT 1403
 DB 506 TTCAATACTCAGCATTTACCATTTGGTTGAGGCCAAATTACAGAGCTTCTCAATCAGATT 565
 QY 1404 TACATCTCCATTTTCATTAACCGGGAAATCCCGGACCACTGAGTGTGTCTTGT 1463
 DB 566 TACATCTCCATTTTCATTAACCGGGAAATCCCGGACCACTGAGTGTGTCTTGT 625
 QY 1464 CACTGAGGTTAGATCTGAACCCAGGGTGTCAACNGTCTCTCAACTCCCACTCTGG 1523
 DB 626 CACTGAGGTTAGATCTGAACCA-GGTGTAAACGCTGCTCTCAATCCCA-CTTGG 683
 QY 1524 GCAGTGAAGATTTCCCTCATTTCTTACCTCTCTTAAAGCTAT 1566
 DB 684 GCAGTGAAGATTTCCCTCATTTCTTACCTCTCTTAAAGCTAT 726

RESULT 9
 CA424992/c 672 bp mRNA linear EST 07-NOV-2002
 LOCUS UI-H-FBI-bee-b-03-0-UI s1 NCI CGAP FBI Homo sapiens CDNA clone
 DEFINITION UI-H-FBI-bee-b-03-0-UI 3', mRNA sequence.
 ACCESSION CA424992
 VERSION CA424992.1 GI:24787718
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 672)
 REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: James Martin
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@iowa.edu
 Seq primer: M13 FORWARD
 POLYA=yes.

FEATURES

source

Location/Qualifiers
 1..672
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-FBI-bee-b-03-0-UI"
 /tissue_type="Cell lines"
 /dev_stage="Adult"
 /lab_host="NCI CGAP FBI"
 /note="Organ: Chondrosarcoma; Vector: pRTT3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP_FBI is a normalized cDNA library derived from a pool of mRNA obtained from 3 cell lines from grade II chondrosarcoma tissues. The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dr primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pRTT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCGTACGGAC. The cell lines were provided by Dr James Martin from the University of Iowa.
 TAG_TISSUE=human grade 2 chondrosarcoma cell line pool
 TAG_LIB=UI-H-FBI

ORIGIN TAG_SEQ=CGCTACGGAC"

Query Match 30.9%; Score 613.8; DB 6; Length 672;
 Best Local Similarity 98.8%; Pred. No. 3.6e-155;
 Matches 649; Conservative 0; Mismatches 4; Indels 4; Gaps 3;

QY 1271 TCACATCATCCCGCTTTTATCTATGATCTATAGGCTCGGAGGTGTGTA--ACAACCA 1328
 DB 672 TCATCATACCCCGCTTTTATCTATGATCTATAGGCTCGGAGGTGTGTAACACACCA 613
 QY 1329 CCCCATCGGTGTCTCTCAATACTGACATTAACATTTGTTAGGCCAATTCAGAGCT 1388
 DB 612 CCCCATCGGTGTCTCTCAATACTGACATTAACATTTGTTAGGCCAATTCAGAGCT 553
 QY 1389 TTCTCAATCAGATTATACATCTCCATTTTCAATTAACGGGGAACAATCCCGAGCCACTG 1448
 DB 552 TTCTCAATCAGATTATACATCTCCATTTTCAATTAACGGGGAACAATCCCGAGCCACTG 493
 QY 1449 AGTGTGTGTGTGTCTCACTGAAGGTAGATGTGAACCCAGGGTGTCACTGCTCTCA 1508
 DB 492 AGTGTGTGTGTGTCTCACTGAAGGTAGATGTGAACCCAGGGTGTCACTGCTCTCA 433
 QY 1509 ACTCCCACTCTCTGGGCACTGAGAGATTTTCCCTCATTTTCACTCTCTTAAGCTATGC 1568
 DB 432 ACTCCCACTCTCTGGGCACTGAGAGATTTTCCCTCATTTTCACTCTCTTAAGCTATGC 373
 QY 1569 ACCCTCTCCCACTCTCTGGGCACTGAGAGATTTTCCCTCATTTTCACTCTCTTAAGCT 1627
 DB 372 ACCCTCTCCCACTCTCTGGGCACTGAGAGATTTTCCCTCATTTTCACTCTCTTAAGCT 313
 QY 1628 CCCATCTGGGATAGGAGACCTTCATCAGCTTAACTCTGGGAATGCTCTGCCCCAG 1687
 DB 312 CCCATCTGGGATAGGAGACCTTCATCAGCTTAACTCTGGGAATGCTCTGCCCCAG 253
 QY 1688 TGACTCTTGTGTCTCTCCCACTGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1747
 DB 252 TGACTCTTGTGTCTCTCCCACTGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 193
 QY 1748 TTAGCAGGGGCTCCCAAGGCAAGTCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1807
 DB 192 TTAGCAGGGGCTCCCAAGGCAAGTCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 133
 QY 1808 CTTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1866
 DB 132 CTTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 73
 QY 1867 CCGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1923
 DB 72 CCGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 16

RESULT 10
 BM996902/c 681 bp mRNA linear EST 26-MAR-2002
 LOCUS UI-H-ED0-awx-c-06-0-UI.s1 NCI CGAP_ED0 Homo sapiens CDNA clone
 DEFINITION IMAGE:5824829 3', mRNA sequence.
 ACCESSION BM996902
 VERSION BM996902.1 GI:19721803
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jose Mercuende
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

FEATURES
 source

CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be found
 through the I.M.A.G.E. Consortium/HLN at: <http://image.llnl.gov>
 Seq primer: M13 FORWARD
 POLYA=yes

Location/Qualifiers

1. 681
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5824829"
 /tissue_type="Chondrosarcoma"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP ED0"
 /note="Organ: Left Pubic Bone; Vector: pT73-Pac
 (Pharmacia) with a modified polylinker; Site 1: EcoR I,
 Site 2: Not I; NCI CGAP_ED0 is a CDNA library containing
 the following tissue(s): Chondrosarcoma cell line C85. The
 library was constructed according to Bonaldo, Lennon and
 Soares, Genome Research, 6:791-806, 1996. First strand
 cDNA synthesis was primed with an oligo-dT primer
 containing a Not I site. Double stranded cDNA was ligated
 to an EcoR I adaptor, digested with Not I, and cloned
 directionally into pT73-Pac vector. The oligonucleotide
 used to prime the synthesis of first-strand cDNA contains
 a library tag sequence that is located between the Not I
 site and the (dT)18 tail. The sequence tag for this
 library is GCTCAAGGCT.
 TAG_TISSUE=chondrosarcoma
 TAG_LIB=UI-H-ED0
 TAG_SEQ=CGCTACGGAC"

ORIGIN

Query Match 30.9%; Score 613.8; DB 5; Length 681;
 Best Local Similarity 98.8%; Pred. No. 3.7e-155;
 Matches 659; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

QY 1260 GCACTAGCTCTCATCATCATCCCGCTTTTATCTATGATCTATAGGCTCGGTGTGTCT 1319
 DB 681 GCACTAGCTCTCATCATCATCCCGCTTTTATCTATGATCTATAGGCTCGGTGTGTCT 622
 QY 1320 AACA-ACACACCCCTTACGCTGTCTTCAATACTAGATTAACATTTGTTAGAGCCCA 1378
 DB 621 AACAACACACCCCTTACGCTGTCTTCAATACTAGATTAACATTTGTTAGAGCCCA 562
 QY 1379 ATTCAAGGCTTTCTCAATCAGATTACATCTCCATTTTCAATTAACGGGGAACAATGCC 1438
 DB 561 ATTCAAGGCTTTCTCAATCAGATTACATCTCCATTTTCAATTAACGGGGAACAATGCC 502
 QY 1439 CGAGCACTGAGTGTGTGTCTTGTCTCACTGAAGTTAGTGAACCCAGGTTCAACN 1498
 DB 501 CGAGCACTGAGTGTGTGTCTTGTCTCACTGAAGTTAGTGAACCCAGGTTCAACN 442
 QY 1499 GGTGCTCTCACTCCCACTCTGGGCACTGAGAGATTTTCCCTCATTTCTACTCTCT 1558
 DB 441 GGTGCTCTCACTCCCACTCTGGGCACTGAGAGATTTTCCCTCATTTCTACTCTCT 382
 QY 1559 AAGGCTATGACCCCTCCCACTCTTCCAGCTGGGGGAT-GGGGGAGTCAATAGAAA 1617
 DB 381 AAGGCTATGACCCCTCCCACTCTTCCAGCTGGGGGATGGGGGAGTCAATAGAAA 322
 QY 1618 GGGCCCATCTCCCATCTGGGATAGGAGACTTCATCAGCTTAACGCTGGGGAATGCCG 1677
 DB 321 GGGCCCATCTCCCATCTGGGATAGGAGACTTCATCAGCTTAACGCTGGGGAATGCCG 262
 QY 1678 CTGCCCCAGTGAATCTTTGTTCTCTCCACATACAGAACAGAGGTGAGAGGAG 1737
 DB 261 CTGCCCCAGTGAATCTTTGTTCTCTCCACATACAGAACAGAGGTGAGAGGAG 202
 QY 1738 TGGGCTCAGTTAGCAGGGGCTCCCAAGGCAAGTCAAGCTCTCTCTCATGCTCTCTG 1797

Db	201	TGGGTCATGTTAGCAGGGGTCGCCAGGGCAAGTCAGCCTCCTCCCTCATGCTCTCTG	142
Qy	1798	GTCAGTGTGCTTTAGGGTGGCCTCTCACTCCCAACCACTCTGGG-CCCTTGGGGAGAGACT	1856
Db	141	GTCAGTGTGCTTTAGGGT-GCCTCTCACTCCCAACCACTCTGGGCCCCCTTGGGGAGAGACT	83
Qy	1857	GGGAGGGGGGCGCTGGGAGACCCCTGAGCCTGGAACTGTATATACATATAAGACACTC	1916
Db	82	GGGGAGGGGGCGCTGGGAGACCCCTGAGCCTGGAACTGTATATACATATAAGAGACTC	23
Qy	1917	TCACAGA	1923
Db	22	TCACAGA	16

RESULT 11	
BU624626/c	
LOCUS	BU624626
DEFINITION	BU624626 676 bp mRNA linear EST 23-SEP-2002
ACCESSION	U1-H-Fg1-bgk-d-20-0-U1 si NC1 CGAP_Fg1 Homo sapiens CDNA clone
VERSION	U1-H-Fg1-bgk-d-20-0-U1 3', mRNA sequence.
KEYWORDS	BU624626
SOURCE	BU624626.1 GI:23290841
ORGANISM	EST.
	Homo sapiens (human)
	Homo sapiens

REFERENCE	1 (bases 1 to 676)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL	Tumor Gene Index
COMMENT	Unpublished (1997)
	Contact: Robert Strausberg, Ph.D.

Unpublished (1997)
Contact: Robert Strusberg, Ph.D.
Email: cgaab5-10@mail.nih.gov
Tissue Procurement: James Martin
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq Primer: W13 FORWARD
POLYA=Yes.

FEATURES	Location/Qualifiers
source	1. .676

```

source
1. .676
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="U1-H-Fgl-bgr-d-20-0-U"
/tissue_type="Cell lines"
/dev_stage="Adult"
/lab_host="MDH103 (Life Technologies)"
/clone_idb="NCI CGAP Fgl"
/note="Organ: Enchondroma; Vector: pRTT3-Pac (Pharmacia)
with a modified polylinker; Site_1: EcoR I; Site_2: Not I;
NCI CGAP Fgl is a normalized cDNA library obtained from a
pool of mRNA from 2 cell lines from Enchondroma tissues.
The library was constructed according to Bonaldo, Lennon
and Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pRTT3-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is CGGTCACTC. The cell lines were provided by Dr.
James Martin from the University of Iowa.
TAG TISUE=Enchondroma cell line (Mix of EN1 and EN2)
TAG_LIB=U1-H-Fgl
TAG_SEQ=CGGTCACTC"

```

ORIGIN	Query Match
--------	-------------

30.9%; Score 613.2; DB 5; Length 676;

Best Local Similarity 99.1%; Pred. No. 5.3e-155;
Matches 637; Conservative 0; Mismatches 4; Indels 2
Gaps 2

Oy	1283	CCCTTTTAATCTAAGAACTCAATAGAGCTCGGTGGTGTGAACAACACAC-CCCTATCGTGT	1341
Db	660	CCCTTTTAATCTAAGAACTCAATAGAGCTCGGTGGTGTGAACAACACACACCTTATCGTGT	601
Oy	1342	CCTTCAAAATCTACGCATTAACATTTGGTTGAGGCGCAATTTCAAGCTTTCTCAATACGA	1401
Db	600	CCTTCAAAATCTACGCATTAACATTTGGTTGAGGCGCAATTTCAAGCTTTCTCAATACGA	541
Oy	1402	TTTAAATCTCCATTTTCATTTAAGGGGAAAACATCCCGAGCCATGAGTGTGCTTT	1461
Db	540	TTTAAATCTCCATTTTCATTTAAGGGGAAAACATCCCGAGCCATGAGTGTGCTTT	481
Oy	1462	GTCATCTGAAGTTGATCTGAACCCAGAGGTGTCAACNGTGTCTTCAATCCCACTCT	1521
Db	480	GTCATCTGAAGTTGATCTGAACCCAGAGGTGTCAACNGTGTCTTCAATCCCACTCT	421
Oy	1522	GGGCACTGAGAGATTTTCCCTCATTTCTACCTCTTAAGGCTATGACCCCTCCACG	1581
Db	420	GGGCACTGAGAGATTTTCCCTCATTTCTACCTCTTAAGGCTATGACCCCTCCACG	361
Oy	1582	TCTTTCAGCTGGGGGATGGGGGGAGTATAGGAAAAGCCCCCATCTCCATCTGGGATG	1641
Db	360	TCTTTCAGCTGGGGGATGGGGGGAGTATAGGAAAAGCCCCCATCTCCATCTGGGATG	301
Oy	1642	GGACCTTCCATCAGCCTTTAACCTCGGGAAAATGCTGTGCCCCAGAGACTCTGTGTTT	1701
Db	300	GGACCTTCCATCAGCCTTTAACCTCGGGAAAATGCTGTGCCCCAGAGACTCTGTGTTT	241
Oy	1702	GTCCTCCACATACGAAGCAGGGGTGGAGGGGAGGGGTCTCATGTTAGCAGGGGTCC	1761
Db	240	GTCCTCCACATACGAAGCAGGGGTGGAGGGGAGGGGTCTCATGTTAGCAGGGGTCC	181
Oy	1762	CAGGGCAAGTCAGGCTCTCCCTCCATGAGCTCTGTGTCAGTGTGCTTGAAGGTGGCTC	1821
Db	180	CAGGGCAAGTCAGGCTCTCCCTCCATGAGCTCTGTGTCAGTGTGCTTGAAGGTGGCTC	121
Oy	1822	TCACTCCACATCTCTGGG-CCCTTGGGAGAGACTGGGAGGGGCGGTGGAGAGCC	1880
Db	120	TCACTCCACATCTCTGGGCGCCCTTGGGAGAGAGCTGGGAGGGGCGGTGGAGAGCC	61
Oy	1881	TGAGCGTGGAACTGTATACCAATTAAGGACATCTCCACAGA	1923
Db	60	TGAGCGTGGAACTGTATACCAATTAAGGAGAGCTCCACAGA	18
RESULT 12			
BF306561			
LOCUS	636 bp	mRNA	linear
DEFINITION	601888988EF1 NIH_MGC_17	Homo sapiens	cdNA clone IMAGE:4122788 5',
ACCESSION	BF306561		
VERSION	BF306561.1	GI:11253663	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	NIH-MGC http://mgc.ncl.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Straubeberg, Ph.D.		

RESULT	12
BFP06561	
LOCUS	
DEFINITION	BFP06561 636 bp mRNA linear EST 21-NOV-2000 6018898986F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4122785,
ACCESSION	
VERSION	BFP06561
KEYWORDS	BFP06561.1 GI:11253663
SOURCE	EST.
ORGANISM	Homo sapiens (human)
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	
AUTHORS	NIH-MGC Http://mgc.nci.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabs+email.nih.gov Tissue procurement: ATCC CDNA library Preparation: Ling Hong/Rubin Laboratory CDNA sequencing by: Inocyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: Image.llnl.gov Plate: LHCMD1006 row: 1 column: 21

QY 431 TCCTAAACCCATTACACAGCTCCAGACATCTCAGCCAGGTTCTAGAGCCGCCAGC 490
 Db 193 TCCTAAACCCATTACACAGCTCCAGACATCTCAGCCAGGTTCTAGAGCCGCCAGC 252
 QY 491 CCCGACCCCGAGGTGACATCTCTCCCAACTAGAGGCTCTCACTCTATCCAGGCAAGC 550
 Db 253 CCCGACCCCGAGGTGACATCTCTCCCAACTAGAGGCTCTCACTCTATCCAGGCAAGC 312
 QY 551 CAGGACCTCCCTGCTGACACATGATGCCAGATTTGAGATTTGGCTCCGTCACCTAA 610
 Db 313 CAGGACCTCCCTGCTGACACATGATGCCAGATTTGAGATTTGGCTCCGTCACCTAA 372
 QY 611 TCAGAGTACAGGGGCTGAGGCTCAGGAGAGAGATTTAAAGAACCCACTGTGGTCAAG 670
 Db 373 TCAGAGTACAGGGGCTGAGGCTCAGGAGAGAGATTTAAAGAACCCACTGTGGTCAAG 432
 QY 671 GGAATGGGACCCAGAGGACATATGAGGAGCTGTCAGACAGACAGACAGACAAACCT 730
 Db 433 GGAATGGGACCCAGAGGACATATGAGGAGCTGTCAGACAGACAGACAGACAAACCT 492
 QY 731 CTGATCTATGAGTCTCTGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 790
 Db 493 CTGATCTATGAGTCTCTGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 552
 QY 791 GGGAGTGGAGAGACAGAGGGAAGGTACAGGCAAGGCTGCTATCTTAAGTGAATTAAT 850
 Db 553 GGGAGTGGAGAGACAGAGGGAAGGTACAGGCAAGGCTGCTATCTTAAGTGAATTAAT 612
 QY 851 TGCCCGAGGCTCAGACAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 892
 Db 613 TGCCCGAGGCTCAGACAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 654

RESULT 14
 BF203719 976 bp mRNA linear EST 06-NOV-2000
 LOCUS 601865380P1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:409458 5',
 DEFINITION mRNA sequence.
 ACCESSION BF203719
 VERSION BF203719.1 GI:11097305
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 976)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: rgs@pds-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
 Plate: LINC967 row: P column: 19
 High quality sequence stop: 676.
 Location/Qualifiers
 1..976
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:409458"
 /tissue_type="rhabdomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NIH_MGC_17"
 /note="Organ: muscle; Vector: pOTB7; Site: 1; Ecort;
 Site: 2; XhoI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by

ORIGIN
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

Query Match 29.6%; Score 587.4; DB 2; Length 976;
 Best Local Similarity 93.1%; Pred. No. 6e-148;
 Matches 712; Conservative 0; Mismatches 43; Indels 10; Gaps 9;

QY 761 GACCAAGGACCTGAGACCTCTTGCCCAAGGGGAGTGGGAGAGACAGAGGGAGGTCA 820
 Db 1 GACCAAGGACCTGAGACCTCTTGCCCAAGGGGAGTGGGAGAGACAGAGGGAGGTCA 59
 QY 821 GGCAGAGGTGCTTATCTTAAGTGAATTAATGCTCCGAGGCTCAGCAAGCCAGAGAG 880
 Db 60 GGCAGAGGTGCTTATCTTAAGTGAATTAATGCTCCGAGGCTCAGCAAGCCAGAGAG 119
 QY 881 ACAGCCGTGACGGTAACTTCCCTCTACAGACCTCCAGCCCAAGCCGACAGAGCAGC 940
 Db 120 ACAGCCGTGACGGTAACTTCCCTCTACAGACCTCCAGCCCAAGCCGACAGAGCAGC 179
 QY 941 TGCCTGCCACCCCGTGCACAGCCAGCTGGCTGTGCAGAGGACAGACCATGACATC 1000
 Db 180 TGCCTGCCACCCCGTGCACAGCCAGCTGGCTGTGCAGAGGACAGACCATGACATC 238
 QY 1001 TGTATATGATGGGCTTTTCCATATACAGCTGTGTTCTGTGAAAACCTGATGAATCT 1060
 Db 239 TGTATATGATGGGCTTTTCCATATACAGCTGTGTTCTGTGAAAACCTGATGAATCT 298
 QY 1061 CGCTCTGCGCTGCTGGGGGCTCCAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 1120
 Db 299 CGCTCTGCGCTGCTGGGGGCTCCAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 358
 QY 1121 CCTTCTCCCTCCACAGGCTGTGTTCTNTGGGCTGTCTCCATGACAGATCACCTA 1180
 Db 359 CCTTCTCCCTCCACAGGCTGTGTTCTNTGGGCTGTCTCCATGACAGATCACCTA 418
 QY 1181 ACAGAGATGAGACAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 1240
 Db 419 ACAGAGATGAGACAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAG 476
 QY 1241 CTGACACCTTCCCTCCACAGGCTGTGTTCTNTGGGCTGTCTCCATGACAGATCACCTA 1299
 Db 477 CTGACACCTTCCCTCCACAGGCTGTGTTCTNTGGGCTGTCTCCATGACAGATCACCTA 536
 QY 1300 TCTATATGCTCGTGTGTGTATACACACACACCTTATGTTGCTTCAATATCTAGCAT 1359
 Db 537 TCTATATGCTCGTGTGTGTATACACACACACCTTATGTTGCTTCAATATCTAGCAT 595
 QY 1360 TACCATTTGTTAGGCGCAATCCCGAGCCACTGAGTGTCTGTGCTTTGTATCTGAATTC 1419
 Db 596 TACCATTTGTTAGGCGC-AATTGAGAGC-TTCTCAATCAATTTACATCTCATTTCA 653
 QY 1420 ATTAAAGGGGAAACATCCCGAGCCACTGAGTGTCTGTGCTTTGTATCTGAATTC 1479
 Db 654 TTAAAGGGGAGACGATCCCGAGCCACTGAGTGTCTGTGCTTTGTATCTGAATTC 711
 QY 1480 TGAACCCAGGGGTGTCACAGGCTGTCTCAACTGCCCACTGAGG 1524
 Db 712 TGAACCCAGGGGTGTCACAGGCTGTCTCAACTGCCCACTGAGG 756

RESULT 15
 BF203383 956 bp mRNA linear EST 06-NOV-2000
 LOCUS 601865935P1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:409646 5',
 DEFINITION mRNA sequence.
 ACCESSION BF203383
 VERSION BF203383.1 GI:11096969
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 956)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabds@mail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (ULNL)
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ULNL at: image.lnl.gov
 Plate: LLC965 row: n column: 23
 High quality sequence stop: 682.
 Location/Qualifiers

FEATURES

1..956

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:409846"
 /issue_type="Thadomysarcoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_1b="NIH_MGC_17"
 /note="Organ: muscle; Vector: pOTB7; Site: 1: EcoRI; Site: 2: XhoI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(C). Size-selected for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Query Match

29.4%; Score 585; DB 2; Length 956;

Best Local Similarity 89.5%; Pred. No. 2.7e-147; Matches 708; Conservative 0; Mismatches 72; Indels 11; Gaps 7;

```

QY 766 GGGACCTGGAGACCTTTGGCCAGGGAGTGGAGACAGAGGGAAGTCAACAGCAA 825
DB 1 GGGACCTGGAGACCTTTGGCCAGGGAGTGGAGACAGAGGGAAGTCAACAGCAA 60
QY 826 GGGTGCTTCTTAAGTGAAGTAATTCGCCAGGGCTCAGCAAGGCCAAGAGAGACAGC 885
DB 61 GGGTGCTTCTTAAGTGAAGTAATTCGCCAGGGCTCAGCAAGGCCAAGAGAGACAGC 120
QY 886 CGTAGAGTAAGTTCCTTACCAAGCTCCAGGCCAGCGACGAGGAGGCTGACT 945
DB 121 CGTAGAGTAAGTTCCTTACCAAGCTCCAGGCCAGCGACGAGGAGGCTGACT 180
QY 946 GCCCAGCCCGTGCCTCCAGCCAGCTGCTGTCAGGGCAGAGGCCATGCCATCTGAT 1005
DB 181 GCCCAGCCCGTGCCTCCAGCCAGCTGCTGTCAGGGCAGAGGCCATGCCATCTGAT 239
QY 1006 ATAGATGGGGTTTTTCCAAATACAGCTGTTGCTGAATACTGATGAATCTGCCGCTC 1065
DB 240 ATAGATGGGGTTTTTCCAAATACAGCTGTTGCTGAATACTGATGAATCTGCCGCTC 299
QY 1066 CTGCGCTGCTGGGGCTTCAGGCAAGGCAAGTGGGGTGGGGCTGATCCTTC 1125
DB 300 CTGCGCTGCTGGGGCTTCAGGCAAGGCAAGTGGGGTGGGGCTGATCCTTC 359
QY 1126 TCCCTCCCAAGGCTGTGTCTCTGATGAGGCTGCTCCATGCAAGATCACTTAACGA 1185
DB 360 TCCCTCCCAAGGCTGTGTCTCTGATGAGGCTGCTCCATGCAAGATCACTTAACGA 419
QY 1186 GATGAGAGCAGGAGCATGATGGGCTTGGGCTCTCGAGTTGAGCCCAAGCTCTTGC 1245
DB 420 GATGAGAGCAGCAGCATGATGGGCTTGGGCTCTCGAGTTGAGCCCAAGCTCTTGC 478
QY 1246 CACTTCCCTCCGCGAGTCAAGCTCTCAATCCATCCCTCTTAATCTATGAATCTATA 1305
DB 479 CACTTCCCTCCGCGAGTCAAGCTCTCAATCCATCCCTCTTAATCTATGAATCTATA 538

```

```

QY 1306 GGGTGGTGTGTGTAAACAACACACCCCTATGTTGCTTCAAAATCTAGCATTAAC 1365
DB 539 GGGTGGTGTGTGTAAACAACACA-CCCTATGTTGCTTCAAAATCTAG-ATTACAT 596
QY 1366 TGGTGAAGCCAAATTCAGAGCTTTCGCAATGATTAAGATCTCCATTTTCAATTAAC 1425
DB 597 TGGTGAAGCCAAATTCAGAGCTTTCGCAATGATTAAGATCTCCATTTTCAATTAAC 656
QY 1426 GGGGAAACATCCCGAGCCACTGAGTGTGCTTTGTCACTGAAGATTAGATCTGAAC 1485
DB 657 GG-----ACATCCGAGCACTAGTG-TGTGCTATGTCAAGTGAAGATTAGATCTG-ACC 709
QY 1486 CAGGTGTCAACGCTGCTCTCAACTCCCACTCGGGCACTGAGAGATTTTCCCTC 1545
DB 710 CAGGTGTCAACGCTGCTCTTAATCTCCCACTCGGGCACTGAGAGATTTTCCCTC 769
QY 1546 ATTCTACCTCT 1556
DB 770 TTTTATGTTT 780

```

Search completed: October 24, 2004, 15:41:05
 Job time : 6400 secs

QY 121 CCTGAAGACATCATTTGTCAACACCTCCAAACAAACCAATCGAGGTGGGGGTGAGAA 180
Db 121 CTTGAAGACATCATTTGTCAACACCTCCAAACAAACCAATCGAGGTGGGGGTGAGAA 180
QY 181 AGCTGCGGTGAAACGGAACCTGTATCATMAAACCTTACGACTACAAAGTGGCAACT 240
Db 181 AGCTGCGGTGAAACGGAACCTGTATCATMAAACCTTACGACTACAAAGTGGCAACT 240
QY 241 GCGGAGGAGCTGAGACCCCGAGGTGGGTGACCTTGGCTACTGTGAGAGGACGACAG 300
Db 241 GCGGAGGAGCTGAGACCCCGAGGTGGGTGACCTTGGCTACTGTGAGAGGACGACAG 300
QY 301 CCTCACTAGTCCGGGACGCTACCCGATACAGAACAGTCCAGACAGCTTCCGAC 360
Db 301 CCTCACTAGTCCGGGACGCTACCCGATACAGAACAGTCCAGACAGCTTCCGAC 360
QY 361 GGAAGATCAAAATTTGAGTGCCTCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCC 420
Db 361 GGAAGATCAAAATTTGAGTGCCTCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCC 420
QY 421 CAGCAAGGCTGCTAACCCCATTAACAAGCTCCAGGACATCTGAGCCAGGTTCTAGC 480
Db 421 CAGCAAGGCTGCTAACCCCATTAACAAGCTCCAGGACATCTGAGCCAGGTTCTAGC 480
QY 481 CCCACGACACCCGACACCCGAGGTGACCATCTTCCAACTAGGGCTCTCACTATC 540
Db 481 CCCACGACACCCGACACCCGAGGTGACCATCTTCCAACTAGGGCTCTCACTATC 540
QY 541 CAGGGAGGACGAGGACCTCCCTGAGCTGACATATGATGCCAGATTTGAGTTGGCTC 600
Db 541 CAGGGAGGACGAGGACCTCCCTGAGCTGACATATGATGCCAGATTTGAGTTGGCTC 600
QY 601 CGTCACTTAATCAGAGTACAGGGGCTGGGGCTCAGGAGGAAAGATCTAAAGAACCACT 660
Db 601 CGTCACTTAATCAGAGTACAGGGGCTGGGGCTCAGGAGGAAAGATCTAAAGAACCACT 660
QY 661 GTGGGTGAGGGGAAATGGGACAGACATATGGGCAAGCTCTGACAGACAGACAGCA 720
Db 661 GTGGGTGAGGGGAAATGGGACAGACATATGGGCAAGCTCTGACAGACAGACAGCA 720
QY 721 GACAAACCTCTGATCTATGAGTCTCTGACAGGACAGGAGACAGGAGCTGGAACCT 780
Db 721 GACAAACCTCTGATCTATGAGTCTCTGACAGGACAGGAGACAGGAGCTGGAACCT 780
QY 781 CTTGGGCAAGGGGAGTGGGAGAGACAGAGGAAAGTCAAGGCAAGGGTGTCTATTAAG 840
Db 781 CTTGGGCAAGGGGAGTGGGAGAGACAGAGGAAAGTCAAGGCAAGGGTGTCTATTAAG 840
QY 841 TGAACCTAATTTGCCGAGGGCTCAGCAAGGCCAAGAGAGAAGCCGTGACGTAACCTT 900
Db 841 TGAACCTAATTTGCCGAGGGCTCAGCAAGGCCAAGAGAGAAGCCGTGACGTAACCTT 900
QY 901 CCCCTCTACAGCTCTCAAGGCCCAAGCCAGAGAGAGGCTGCTGACACCCGTGCC 960
Db 901 CCCCTCTACAGCTCTCAAGGCCCAAGCCAGAGAGAGGCTGCTGACACCCGTGCC 960
QY 961 CCAGCAGCTGCTGTGACAGGGGACAGACATCCACATCTGTATATGATGGGTTTTT 1020
Db 961 CCAGCAGCTGCTGTGACAGGGGACAGACATCCACATCTGTATATGATGGGTTTTT 1020
QY 1021 CCAATACAGCTGTGCTGTGAAAACTGATGAACTCTGCGCTGCTGCGCTGCTGCGG 1080
Db 1021 CCAATACAGCTGTGCTGTGAAAACTGATGAACTCTGCGCTGCTGCGCTGCTGCGG 1080
QY 1081 CCTCAAGGACAGGACCAAGTGGGGTTGGGGGTGGGGCTGCTCTCTCTCCCAAGGCC 1140
Db 1081 CCTCAAGGACAGGACCAAGTGGGGTTGGGGGTGGGGCTGCTCTCTCTCCCAAGGCC 1140
QY 1141 TGTGTTCTGAGGCTGCTCCCATGACAGAGATACCTTAACAGATGAAAGCAAGGCC 1200
Db 1141 TGTGTTCTGAGGCTGCTCCCATGACAGAGATACCTTAACAGATGAAAGCAAGGCC 1200

QY 1201 ATGATGGGGCTTTGGGTCTCGAGGTGGAGCCCAAGCTTCTTCCACCTCCCTCCGG 1260
Db 1201 ATGATGGGGCTTTGGGTCTCGAGGTGGAGCCCAAGCTTCTTCCACCTCCCTCCGG 1260
QY 1261 CAGTCAAGCTTCCATCCATCCCTCTTATCTATGAACTATAGCTGAGTGTGTA 1320
Db 1261 CAGTCAAGCTTCCATCCATCCCTCTTATCTATGAACTATAGCTGAGTGTGTA 1320
QY 1321 ACAACACACCCCTATGTTGTCTTCAAAATCTACAGATTTACATTTGTTGAGCCAAAT 1380
Db 1321 ACAACACACCCCTATGTTGTCTTCAAAATCTACAGATTTACATTTGTTGAGCCAAAT 1380
QY 1381 TCAGAGCTTTCTCAAAATCAGATTTAACAATCTCCATTTGATTAACGGGAAACATCCCG 1440
Db 1381 TCAGAGCTTTCTCAAAATCAGATTTAACAATCTCCATTTGATTAACGGGAAACATCCCG 1440
QY 1441 AGCCACTGAGTGTCTGCTTGTCTTCACTGAAAGTTAATCTGAAACCAAGGTGTCAAACGC 1500
Db 1441 AGCCACTGAGTGTCTGCTTGTCTTCACTGAAAGTTAATCTGAAACCAAGGTGTCAAACGC 1500
QY 1501 TGCTCTCAATCCCACTCTGAGGACCTGAGAGTATTTCCCTCATTTCACTCTCTTA 1560
Db 1501 TGCTCTCAATCCCACTCTGAGGACCTGAGAGTATTTCCCTCATTTCACTCTCTTA 1560
QY 1561 GGCATATGACCCCTCCCACTCTTCCAGCTGAGGAGTGGGGAGTCAAGGAAAGGC 1620
Db 1561 GGCATATGACCCCTCCCACTCTTCCAGCTGAGGAGTGGGGAGTCAAGGAAAGGC 1620
QY 1621 CCCATCTCCATTTGGGATAGGGAACCTTCCATCAAGCTTAACTCTGGGAAATGCTGCTG 1680
Db 1621 CCCATCTCCATTTGGGATAGGGAACCTTCCATCAAGCTTAACTCTGGGAAATGCTGCTG 1680
QY 1681 CCCCCAGTACTCTTGTGTTGCTTCCACATACAGAAACAGGAGGAGGAGGAGTGG 1740
Db 1681 CCCCCAGTACTCTTGTGTTGCTTCCACATACAGAAACAGGAGGAGGAGGAGTGG 1740
QY 1741 GTCTCAGTTAGCAGGGGTCCTCCAGGACAGTCAAGCTCTCCCTCAATGCTCTGTC 1800
Db 1741 GTCTCAGTTAGCAGGGGTCCTCCAGGACAGTCAAGCTCTCCCTCAATGCTCTGTC 1800
QY 1801 AGTGTCCCTTAAAGGTGGCTCTCACTCCACCACTCTGGGACCTTGGGGAGAGACTGGG 1860
Db 1801 AGTGTCCCTTAAAGGTGGCTCTCACTCCACCACTCTGGGACCTTGGGGAGAGACTGGG 1860
QY 1861 AGGGGCTCGTGGAGAGCCCTGACGCTGGAACCTGTATCACAATTAAGAGACAGTCTCAC 1920
Db 1861 AGGGGCTCGTGGAGAGCCCTGACGCTGGAACCTGTATCACAATTAAGAGACAGTCTCAC 1920
QY 1921 AGACNTCTGAGAGGCGCTGCCNGAGTCTCAAACTTGAAGGACAGGCGNNNCTTACTTG 1980
Db 1921 AGACNTCTGAGAGGCGCTGCCNGAGTCTCAAACTTGAAGGACAGGCGNNNCTTACTTG 1980
QY 1981 AGAGAAA 1987
Db 1981 AGAGAAA 1987

RESULT 2
AX017502 2192 bp DNA linear PAT 07-SEP-2000
LOCUS AX017502
DEFINITION Sequence 48 from Patent WO9947655.
ACCESSION AX017502
VERSION AX017502.1 GI:10042299
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Schmitt,A., Specht,T., Dahl,B., Hinzmann,B., Rosenthal,A. and
Pilarsky,C.
TITLE Human nucleic acid sequences from normal breast tissue
JOURNAL Patent: WO 9947655-A 48 23-SEP-1999;

SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
Location/Qualifiers
1. .2192

FEATURES
source /organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 84.8%; Score 1684.4; DB 6; Length 2192;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 1859; Conservative 0; Mismatches 73; Indels 24; Gaps 11;

33 GCCCATCCCAACCAAGCCCGGTGGGCGAGCAACATCAAGACCTTGAAGAGCGCTATGA 92
258 GGCCTTCCAGCCCGCCCGGTGGGCGAGCAACATCAAGACCTTGAAGAGCGCTATGA 317
93 GTTTGGGTGAGTGAAGTGAAGTTCACCTGAAGACATCATGTGACCAACCTCCAAAGA 152
318 GTTTGGGTGAGTGAAGTGAAGTTCACCTGAAGACATCATGTGACCAACCTCCAAAGA 374
153 AACCAACATGAGGTGGGCGGTGAAGAAAGCTGCGGCTGAACGGAACATGTAATGAA 212
375 CAACCAATGAGGTGGGCGGTGAAGAAAGCTGCGGCTGAACGGAACATGTAATGAA 432
213 CACCTTAACGACTACACAAAGTGCAGACTGCGGAGACGTGAGACCCCGAGCGGTGA 272
433 TTGGCT-----CACAAAGTGCAG--CTGCGGAGGACGTGAGACCC--GACGTGGTGA 481
273 CCTCGGCTACTGTGCGGAGAGCGGACAGCTCTCACTGTCGCGGACAGCGTCAACCCGATA 332
482 CTTGG---CTGCGGAGGAGAGCGGACAGCTCTCACTGTCGCGGACAGCGGTCAACCCGATA 538
333 CAGAACAGTTCAGAGAGACTTTCGGAAGGAGATCAAAATCTGAAGTCTCTCTCTCC 392
539 CAGAACAGTTCAGAGAGACTTTCGGAAGGAGATCAAAATCTGAAGTCTCTCTCTCC 598
393 TTTTCCCTGT--CCCCCGCCCGACGCTGCGGCAAAAGCTGCTGAACCCCATTAACAAG 451
599 TTTTCCCTGTGCCCCCGCCCGACGCTGCGGCAAAAGCTGCTGAACCCCATTAACAAG 658
452 CTCAGAGACATCTCAGCCAGGTTTGAAGCCCAAGCAAGCCCAAGCCAGGTGAACAT 511
659 CTCAGAGACATCTCAGCCAGGTTTGAAGCCCAAGCAAGCCCAAGCCAGGTGAACAT 718
512 CTTCCCAACTGAAGGCGCTCACTTAACAGGCGAGGCGAGGAGCTCCCTGGCTGACA 571
719 CTTCCCAACTGAAGGCGCTCACTTAACAGGCGAGGCGAGGAGCTCCCTGGCTGACA 778
572 CATGATCCCAAGATTTGAGATTTGGCTCCGTCACTTAATCCAGAGTACAGGGGCTGG 631
779 CATGATCCCAAGATTTGAGATTTGGCTCCGTCACTTAATCCAGAGTACAGGGGCTGG 838
632 TCAGGGAAGGAATCTTAAGAACCCCACTGGGTCAAGGGATGGGACCAAGAGACAT 691
839 TCAGGGAAGGAATCTTAAGAACCCCACTGGGTCAAGGGATGGGACCAAGAGACAT 898
692 ATGGGCAAGCTCTCAGAGACAGACAGACAAACCTCTGATCTTAAGAGTCTGCA 751
899 ATGGGCAAGCTCTCAGAGACAGACAGACAAACCTCTGATCTTAAGAGTCTGCA 958
752 GGGCAAGGGGACAGGGAACCTGGAACCTCTTGGCCAAAGGAGTGGGAGAGACAAGG 811
959 GGGCAAGGGGACAGGGAACCTGGAACCTCTTGGCCAAAGGAGTGGGAGAGACAAGG 1018
812 AAGGTCAAGGCAAGGCTATCTAAGTGAACCTAATTGCCGAGGCTCAGCAAGG 871
1019 AAGGTCAAGGCAAGGCTATCTAAGTGAACCTAATTGCCGAGGCTCAGCAAGG 1078
872 CAAGAGAGACAGCGGTGACGGTAACTTCCTCTTACAGCTCCAAAGCCCAAGCG 931
1079 CAAGAGAGACAGCGGTGACGGTAACTTCCTCTTACAGCTCCAAAGCCCAAGCG 1138

922 CGAGCAGGCTCCTGCGCCACCCCGTCCCGACAGCAGTGGCTGTGCCAGGCGAGACCA 991
1139 CGAGCAGGCTCCTGCGCCACCCCGTCCCGACAGCAGTGGCTGTGCCAGGCGAGACCA 1198
992 TGCCCATCTGTAATATGATGAGGAGGTTTTCGAATACAGCTGGTTCGTGAATAAATGCA 1051
1199 TGCCCATCTGTAATATGATGAGGAGGTTTTCGAATACAGCTGGTTCGTGAATAAATGCA 1258
1052 AAACCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 1111
1259 AAACCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 1318
1112 GGGGTGCTGCTTCTCCCTCCCAAGAGGCTGTTTNTGAGGCGTGTCCCATGAGACAG 1171
1319 GGGGTGCTGCTTCTCCCTCCCAAGAGGCTGTTTNTGAGGCGTGTCCCATGAGACAG 1378
1172 GATCACTAAGAGATGGAAGCAAGGACATGATGAGGCTTTGAGTCTTCAGAGTTGGA 1231
1379 GATCACTAAGAGATGGAAGCAAGGACATGATGAGGCTTTGAGTCTTCAGAGTTGGA 1438
1232 CCCCAGCTTCTTCCCACTTCCCTCC--GGCAGTCACTTCCATCCATCCCTCTTTA 1290
1439 CCCCAGCTTCTTCCCACTTCCCTCC--GGCAGTCACTTCCATCCATCCCTCTTTA 1498
1291 ATCTATGATCTATGAGGCTGAGTGTGTGAACA--ACACACCCCTATCGTGTCTTCAA 1349
1499 ATCTATGATCTATGAGGCTGAGTGTGTGAACA--ACACACCCCTATCGTGTCTTCAA 1558
1350 TACTAGCATTAACATTTGTTGAGGCGAATTCAGAGCTTTCTCAATCAATTTTACAT 1409
1559 TACTAGCATTAACATTTGTTGAGGCGAATTCAGAGCTTTCTCAATCAATTTTACAT 1618
1410 CTCGATTTTCAATTAAGGAGGAAACATCCCGAGGACATGAGTGTGCTTTGTACATGA 1469
1619 CTCGATTTTCAATTAAGGAGGAAACATCCCGAGGACATGAGTGTGCTTTGTACATGA 1678
1470 AGGTATGATCTGAAACCCAGGAGTGTCAACNGTGTCTCTCACTCCACCTGTGGCACTG 1529
1679 AGGTATGATCTGAAACCCAGGAGTGTCAACNGTGTCTCTCACTCCACCTGTGGCACTG 1738
1530 AGGATATTTTCCCTCATTTTACCTCTTAAGGCTATGCAACCTCCCAAGCTTCCAG 1589
1739 AGGATATTTTCCCTCATTTTACCTCTTAAGGCTATGCAACCTCCCAAGCTTCCAG 1798
1590 CTGGGAGATGGGAGAGCATAGGAAGACCCCATCTCCATCTGGATGAGGAGCTTC 1649
1799 CTGGGAGATGGGAGAGCATAGGAAGACCCCATCTCCATCTGGATGAGGAGCTTC 1858
1650 CATCAGCCTTAAACCTGAGAAATGCTGTGCTGCTCCCACTGATCTTGTGCTCCCA 1709
1859 CATCAGCCTTAAACCTGAGAAATGCTGTGCTGCTCCCACTGATCTTGTGCTCCCA 1918
1710 CATACAGAGAGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1769
1919 CATACAGAGAGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1978
1770 GTCAAGCTTCTCCCTCAATGCTCTCTGATCACTGATGCTTAAAGGTGCTCTCACTCC 1829
1979 GTCAAGCTTCTCCCTCAATGCTCTCTGATCACTGATGCTTAAAGGTGCTCTCACTCC 2038
1830 ACCACTCTGGG--CCCTTGGGAGAGACCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1888
2039 ACCACTCTGGG--CCCTTGGGAGAGACCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2095
1889 GAACCTGTATACACATTAAGAGACAGTCTCAAGACATCTGAGAGCGCTGCGGAGT 1948
2099 GAACCTGTATACACATTAAGAGACAGTCTCAAGACATTAAGAGAGGCGCTGCGGAGT 2157
1949 TCTCAAACTTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1983
2158 TCTCAAACTTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2192

RESULT 3
BD135192 2192 bp DNA linear PAT 18-SEP-2002
LOCUS Human nucleic acid sequence originating in normal mammary tissue.
DEFINITION
BD135192
ACCESSION BD135192.1 GI:23230137
VERSION JP 2002506639-A/39.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2192)
Speft,T., Hintzman,B., Shcmilt,A., Pirarski,C., Duhl,E. and
Rosenthal,A.
Human nucleic acid sequence originating in normal mammary tissue
Patient: JP 2002506639-A 39 05-MAR-2002;
JOURNAL METAGEN GESELTSCHAFT FUER GENOME FORSCHUNG MBH
COMMENT
CS Homo sapiens (human)
FN JP 2002506639-A/39
PD 05-MAR-2002
PF 19-MAR-1999 JP 2000536838
PI 20-MAR-1998 DE 198 13 835 0
PI THOMAS SPEFT, BERND HINTZMAN, ARMIN SHCMILT, CHRISTIAN PIRARSKI,
PI EDGAR DUHL,
PI ANDRE ROSENTHAL
PC C12N15/09,A61K48/00,A61P35/00,A61P43/00,A61P43/00,C07K14/47,
PC C07K16/18,
PC C12N1/21,C12N5/10,C12P21/02,C12Q1/68,G01N33/68//A61K38/00, PC
C12N15/00,
PC C12N5/00,A61K37/02
CC Human nucleic acid sequence originating in normal mammary
tissue
FH Key Location/Qualifiers
FT source 1..2192
FT /organism='Homo sapiens (human)'.
1..2192
Location/Qualifiers
/organism='Homo sapiens'
/mol_type='Genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Query Match 84.8%; Score 1684.4; DB 6; Length 2192;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 1858; Conservative 0; Mismatches 73; Indels 24; Gaps 11;
QY 33 GCCCATCCCAACCAACCCCGGTGGGCAAGCAATCAAGACCTTAGAGACGCTATGA 92
DB 258 GGCCTTCCCAACCCCGGTGGGCAAGCAATCAAGACCTTAGAGACGCTATGA 317
QY 93 GTTGGCGTGAAGTGAAGACTTCTCACTGAAGATCATGTGACCACTCCCAACA 152
DB 318 GTTGGCGTGAAGTGAAGACTTCTCACTGAAGATCATGTGACCACTCCCAACA 374
QY 153 AACCAACATGAGTGCAGGAGCTGAGAAAGCTGGGCTGAAAGGACACTGATGAAA 212
DB 375 CAACCAATGAGTGCAGGAGCTGAGAAAGCTGGGCTGAAAGGACACTGATGAAA 432
QY 213 CACCTTAGACATCAACAAATGCGGAGCTGCGGAGAGAGTGAACCCGAGCGTGGTGA 272
DB 433 TTTCGCT-----CACAAGTGCAG-CTGCGGAGAGAGTGAACCC--GACGTGCTGA 481
QY 273 CCTCGGCTAGTGCAGGAGAGAGCAAGCTCACTAGTCCGGGACGCGTCAACCGCAT 332
DB 482 CTTGG-----CTTGGGAGAGAGAGAGCTCACTAGTCCGGGACGCGTCAACCGCAT 538
QY 333 CAGAAACGCTCCAGAGACTTTCGAGAGAGATTAATAATCTAGTGCCTTCCCTTCC 392
DB 539 CAGAAACGCTCCAGAGACTTTCGAGAGAGATTAATAATCTAGTGCCTTCCCTTCC 598
QY 393 TTTCCTGT-CCCCCGGCGGAGGCTGCGAGCAAGAGCTGCTTAACCCATTCAACAG 451
|||||

DB 599 TTTCCTGTCCCCCGGCGGAGGCTGCGAGCAAGAGCTGCTTAACCCATTCAACAG 658
QY 452 CTCAGAGACATCTACGCCAGGTTTACGCCCCCAGCACCAGACCCAGAGGACAT 511
DB 659 CTCAGAGACATCTACGCCAGGTTTACGCCCCCAGCACCAGACCCAGAGGACAT 718
QY 512 CTTCCAAACTAGAGGCTTCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 571
DB 719 CTTCCAAACTAGAGGCTTCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 778
QY 572 CATGATGCCCAATTTCAATTTGGCTCCGTACTTAATCCAGATTAAGAGGAGGAG 631
DB 779 CATGATGCCCAATTTCAATTTGGCTCCGTACTTAATCCAGATTAAGAGGAGGAG 838
QY 632 TCAGGAGAGAAATCTTAAGAACCACTGTGGGTGAGGAGGAGGAGGAGGAGGAGGAG 691
DB 839 TCAGGAGAGAAATCTTAAGAACCACTGTGGGTGAGGAGGAGGAGGAGGAGGAGGAG 898
QY 692 ATGGGCAAGCTCTGACAGAGACAGACAGACAGACAAACCTCTGTATTAAGTCTTGA 751
DB 899 ATGGGCAAGCTCTGACAGAGACAGACAGACAGACAAACCTCTGTATTAAGTCTTGA 958
QY 752 GGGCAAGGGGACAGAGGACCTGGAACCTTTGGCCAGAGGGAGTGGAGAGACAGAGG 811
DB 959 GGGCAAGGGGACAGAGGACCTGGAACCTTTGGCCAGAGGGAGTGGAGAGACAGAGG 1018
QY 812 AAGTCAACAGCAAGGAGTCTATCTAAGTGAATTAATTTGCCAGGAGGCTCAAGGCT 871
DB 1019 AAGTCAACAGCAAGGAGTCTATCTAAGTGAATTAATTTGCCAGGAGGCTCAAGGCT 1078
QY 872 CAAGAGAGACAGCGGTGAACGTAATCTCCCTCAACAGGCTCAACGCCCAAGCCAG 931
DB 1079 CAAGAGAGACAGCGGTGAACGTAATCTCCCTCAACAGGCTCAACGCCCAAGCCAG 1138
QY 932 CGAGCAGGCTGCTGCGCAACCCCGTCCAGCCAGCAGGCTGCTGCTGCTGCTGCTGCT 991
DB 1139 CGAGCAGGCTGCTGCGCAACCCCGTCCAGCCAGCAGGCTGCTGCTGCTGCTGCTGCT 1198
QY 992 TGGCAATCTGATATAGATGGGCTTTTCCAAATAGAGTGTGTGTAATAATGCAAG 1051
DB 1199 TGGCAATCTGATATAGATGGGCTTTTCCAAATAGAGTGTGTGTAATAATGCAAG 1258
QY 1052 AAATCCTGCGCTCTGCGCTGCTGCGCTGCTGCGCTGCTGCGCTGCTGCGCTGCTGCT 1111
DB 1259 AAATCCTGCGCTCTGCGCTGCTGCGCTGCTGCGCTGCTGCGCTGCTGCGCTGCTGCT 1318
QY 1112 GGGGCTGCTCTTCT 1171
DB 1319 GGGGCTGCTCTTCT 1378
QY 1172 GATCACTTAACAGAGATGGAAGCAGGAGTGAATGGGCTTTGGGCTCTCGAGGTGA 1231
DB 1379 GATCACTTAACAGAGATGGAAGCAGGAGTGAATGGGCTTTGGGCTCTCGAGGTGA 1438
QY 1232 CCCAGCTTTGCGACTTCCCTCTCC-GGCACTGACTTCTCATCATCCCTCTCTTTA 1290
DB 1439 CCCAGCTTTGCGACTTCCCTCTCCGCGAGTCACTCTCATCATCCCTCTCTTTA 1498
QY 1291 ATCTATGAATCTATAGGCTCGGTGTGAAGA-ACAGCCCTTACGTTGCTTCAA 1349
DB 1499 ATCTATGAATCTATAGGCTCGGTGTGAAGA-ACAGCCCTTACGTTGCTTCAA 1558
QY 1350 TACTCAGATTAACATTTGTTGAGGCAATTCAGAGCTTTCTCAATCAGATTACAT 1409
DB 1559 TACTCAGATTAACATTTGTTGAGGCAATTCAGAGCTTTCTCAATCAGATTACAT 1618
QY 1410 CTCATTTTCTTAAGAGGAGAAACATCCCGGAGCCATGAGTGTGTGTCTCACTGA 1469
DB 1619 CTCATTTTCTTAAGAGGAGAAACATCCCGGAGCCATGAGTGTGTGTCTCACTGA 1678
QY 1470 AGGTTAGATCTGAACCCAGGAGTCAACNGTGTCTCACTCCCACTCTGGGACACTG 1529
DB 1679 AGGTTAGATCTGAACCCAGGAGTCAACNGTGTCTCACTCCCACTCTGGGACACTG 1738

QY	1530	AGGAGTATTTTCCCTCTTCTTACCTCTCTAAGGCTATGACCCCTCCCAAGCTTCCAG	1589
DB	1739	AGAGATATTTTCCCTCTTCTTACCTCTCTAAGGCTATGACCCCTCCCAAGCTTCCAG	1798
QY	1590	CTGGGGATGGGGGAGTCAATAGGAAAAGCCCCATCTCCCATCTGGGATAGGGAACCTTC	1649
DB	1799	CTGGGGATGGGGGAGTCAATAGGAAAAGCCCCATCTCCCATCTGGGATAGGGAACCTTC	1858
QY	1650	CATCAGCTTAACTCTGGGAAATGCTGCTGCTCCCAAGTACATCTTGGTTTCTCTCCA	1709
DB	1859	CATCAGCTTAACTCTGGGAAATGCTGCTGCTCCCAAGTACATCTTGGTTTCTCTCCA	1918
QY	1710	CATCAGAGAGAGGAGTGGAGGAGGAGGAGTCTCAGTTAGAGAGGAGTCCCAAGGACAA	1769
DB	1919	CATCAGAGAGAGGAGTGGAGGAGGAGGAGTCTCAGTTAGAGAGGAGTCCCAAGGACAA	1978
QY	1770	GTCAGCT	1829
DB	1979	GTCAGCT	2038
QY	1830	ACCACTCTGGG-CCTTGGGGGAGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1888
DB	2039	ACCACTCTGGG-CCTTGGGGGAGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2098
QY	1889	GAACCTGTATACACATTAAGAGACAGTCTCAGACAGACNTCTGAGAGCCGCTGCGAGAGT	1948
DB	2099	GAACCTGTATACACATTAAGAGACAGTCTCAGACAGACNTCTGAGAGCCGCTGCGAGAGT	2157
QY	1949	TCTCAACTTATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1983
DB	2158	TCTCAACTTATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	2192
RESULT 4			
LOCUS	CQ721116	2138 bp	DNA
DEFINITION	Sequence 7050 from Patent WO02068579.	linear	PAT 03-FEB-2004
ACCESSION	CQ721116		
VERSION	CQ721116.1	GI:4281973	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE			
AUTHORS	Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.		
TITLE	Kits, such as nucleic acid arrays, comprising a majority of		
	humanexons or transcripts, for detecting expression and other uses		
JOURNAL	Patent: WO 02068579-A 7050 06-SEP-2002;		
FEATURES	Location/Qualifiers		
source	1..2138		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	83.9%; Score 1666.8; DB 6; Length 2138;		
Best Local Similarity	95.7%; Pred. No. 0;		
Matches 1813; Conservative	0; Mismatches 59; Indels 22; Gaps 9;		
QY	33	GCCATCCCAACCAAGCCCGGTGGAGGAGACATCAAGCCCTTAGAGAGCCCTTGA	92
DB	264	GACCTTCCAGCCCGCCCGGTGGAGGAGACATCAAGCCCTTAGAGAGCCCTTGA	323
QY	93	GTTGGGAGTGAAGTGAAGATCTTCACTGAGACATCATTTGTCACCACTTCCAAACA	152
DB	324	GTTGGGAGTGAAGTGAAGATCTTCACTGAGACATCATTTGTCACCACTTCCAAACA	152
QY	153	AACCAATCGAGGTGGGGGTGAAGAACTGGCGGTGAACGACACTTAATGAA	212
DB	1445	CCAGCTTCTTGCACTTCCCTCCGAGGAGTCACTTCAATCATCCCTCTTAA	1504
DB	381	CAACACATCGAGGTGGGG--CTGAGAGCTGGCGGTGAACGACACCGTCAATGAACCC	438
QY	213	CACCTTAGACATCAACAGTCCAGACCTGGGAGAGACATGAGACCCGAGCGGTGA	272
DB	439	TTTGCT-----CAAGTGGCAG-CTGCCGAGAGCTGAGCC--GACGTGGTGA	487
QY	273	CCTCGCTACTGTGGGAGAGAGACAGCTCACTAGTCCGAGGACGGTCAACCCGATYA	332
DB	488	CCTGG---CTCTGGGAGAGAGGAGCCTCACTATCCGGGACAGGGGTCAACCCGATYA	544
QY	333	CAGAACAGCTCAGAGACCTTCCGAGAGGAGATCAAAATCTGAGTGGCTCTCCCTCC	392
DB	545	CAGAACAGCTCAGAGACCTTCCGAGAGGAGATCAAAATCTGAGTGGCTCTCCCTCC	604
QY	393	TTTCCCTGCCCCCGCCCAAGCCTGACAGCAAAAGCTGCTAAACCCATTACAGAC	452
DB	605	TTTCCCTGCCCCCGCCCAAGCCTGACAGCAAAAGCTGCTAAACCCATTACAGAC	664
QY	453	TCAGAGACATTCAGCCAGGTTCTTAGCCCTCAGACACCCAGACCCAGTGAACATC	512
DB	665	TCAGAGACATTCAGCCAGGTTCTTAGCCCTCAGACACCCAGACCCAGTGAACATC	724
QY	513	CTCCCAACTAGGAGCCCTCCACTATCCAGGAGGAGGAGGAGTCCCTGAGCTGACAC	572
DB	725	CTCCCAACTAGGAGCCCTCCACTATCCAGGAGGAGGAGGAGTCCCTGAGCTGACAC	784
QY	573	ATGATGCCAGATTTCAATTTGGCTCCGTCACTTAATCAAGATACAGGGGCTGGGCT	632
DB	785	ATGATGCCAGATTTCAATTTGGCTCCGTCACTTAATCAAGATACAGGGGCTGGGCT	844
QY	633	CAGGAGAGAGATTTAAAGAACCCAGTGGGTAGGGGAAATGGGACAGAGGACATA	692
DB	845	CAGGAGAGAGATTTAAAGAACCCAGTGGGTAGGGGAAATGGGACAGAGGACATA	904
QY	693	TGGGCAAGCTCTGACAGACAGACAGACAGACAACTCTGATCTATGAGTCTTGACG	752
DB	905	TGGGCAAGCTCTGACAGACAGACAGACAGACAACTCTGATCTATGAGTCTTGACG	964
QY	753	GCGAAGGAGACAGGACCTTGAACCTCTTGGCCAAAGGAGGTGGAGAGACAGAGGA	812
DB	965	GCGAAGGAGACAGGACCTTGAACCTCTTGGCCAAAGGAGGTGGAGAGACAGAGGA	1024
QY	813	AGGTCAAGGAGGAGGCTCTATCTAAGTGAATTAATTCAGGAGGCTCAGCAAGGCT	872
DB	1025	AGGTCAAGGAGGAGGCTCTATCTAAGTGAATTAATTCAGGAGGCTCAGCAAGGCT	1084
QY	873	AAGAGAGACAGCCGTGACGTTAACTTCCCTCTTACAGGCTCCAGGCCACGCCACG	932
DB	1085	AAGAGAGACAGCCGTGACGTTAACTTCCCTCTTACAGGCTCCAGGCCACGCCACG	1144
QY	933	GAGCAGGCTGCTGCGCCACCCCGTGGCCCAAGCAGCTGGGCTGACAGGAGGACAT	992
DB	1145	GAGCAGGCTGCTGCGCCACCCCGTGGCCCAAGCAGCTGGGCTGACAGGAGGACAT	1204
QY	993	GCCACATCTGATATAGATGGGATTTTCCAAATACAGCTGTTCTGTGAAAACTGCATGA	1052
DB	1205	GCCACATCTGATATAGATGGGATTTTCCAAATACAGCTGTTCTGTGAAAACTGCATGA	1264
QY	1053	AACCTGCGGCTGCTGCGCTCTGCGGCTCCAGGACAGGACAGTGGGGTTGGGGGG	1112
DB	1265	AACCTGCGGCTGCTGCGCTCTGCGGCTCCAGGACAGGACAGTGGGGTTGGGGGG	1324
QY	1113	GAGCTGCTCTCTCCCTCCACAGGCTGTGTTGTTGAGGCTGCTCCCATCAGACAGG	1172
DB	1325	GAGCTGCTCTCTCCCTCCACAGGCTGTGTTGTTGAGGCTGCTCCCATCAGACAGG	1384
QY	1173	ATCACTTAACAGATGAAAGCAGGAGCATGATGGGCTTTGGGTCTCGAGTTTGAC	1232
DB	1385	ATCACTTAACAGATGAAAGCAGGAGCATGATGGGCTTTGGGTCTCGAGTTTGAC	1444
QY	1233	CCAGCTTCTGACACTTCCCTCC--GGCAGTCACTCTCATCATCCCTCTTTAA	1291
DB	1445	CCAGCTTCTGACACTTCCCTCC--GGCAGTCACTCTCATCATCCCTCTTTAA	1504

TITLE
JOURNAL
PUBLISHED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
REMARK

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimmwood, J., Schmutz, V., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Snailus, D.E.,
 Schercher, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16699-16903 (2002)
 12477932
 2 (bases 1 to 2589)
 Strausberg, R.
 Direct Submission
 Submitted (09-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 On Aug 19, 2003 this sequence version replaced gi:13623438.
 Contact: MGC help desk
 Email: cgaps-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLU)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@nhgri.nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Khong, P., Laric, P., Legaspi, R.,
 Mauro, Q.L., Maiello, C., Makkeri, B., Mascrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Skattcrup, S., Thomas, P.J., Touchman, J.W.,
 Tsurgoun, C., Vogt, J.T., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

FEATURES
source

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNLU at: <http://image.lnl.gov>
 Series: IRAL Plate; 17 Row; 1 Column; 17
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 20070250.
 Location/Qualifiers

1. 2589
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:12642 IMAGE:4131506"
 /tissue_type="Muscle, Thadomyosarcoma"
 /clone_lib="NIH_MGC_17"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 1. 2589
 /gene="HSPB7"
 /note="synonym: CVHSP"
 /db_xref="LocustID:27129"
 494..1006
 /gene="HSPB7"
 /codon_start=1
 /product="heat shock 27kda protein family, member 7
 (cardiovascular)"
 /protein_id="AA006319.1"
 /db_xref="GI:13623439"
 /db_xref="LocustID:27129"
 /translation="MSHRTSTFRAERFSHSSSSSSSSSSSSASRALPAQDPPEKA
 LSFSDGSPMRPHSEPLAPFARGAANIITLDAEFAVDVDFSEDIITTSN
 NHLEVAETLADGTYMNTFAKCOLPEVDVDTSTVSLAREGSLITIRARRPHTEHV
 QQTFREHETI"
 ORIGIN

Query Match 83.7%; Score 1663.6; DB 9; Length 2589;
 Best Local Similarity 95.6%; Pred. No. 0;
 Matches 1811; Conservative 0; Mismatches 61; Indels 22; Gaps 9;

QY 33 GCCCATCCCAAGAGCCCGTGGGAGGCAACATCAAGACCTTAGAGAGCGCTATATA 92
 Db 682 GGGCTTCCAGCCCGCCGCTGGGAGGCAACATCAAGACCTTAGAGAGCGCTATATA 741
 QY 93 GTTGGGGTGGAGTGAAGACTTCTCACTGAAGACATCATTTGTACCAACCTCCAAAGA 152
 Db 742 GTTTGGGGTGGAGTGAAGACTTCTCACTGAAGACATCATTTGTACCAACCTCCAAAGA 798
 QY 153 AACCAATCATGAGAGTGGGGCTGAGAAAGCTGGCGCTGAAAGGCACTGTACATGAAA 212
 Db 799 CAACCAATCATGAGAGTGGGG--CTGAGAAAGCTGGCGCTGACCGGACCTGATGAACAC 856
 QY 213 CACCTTACAGACTACAAAGTCCAGACTGGCGGAGAGACCTGGAGCCCGAGCGGTGAGTA 272
 Db 857 TTGGCT-----CACAAGTGCAG--CTGCGGAGGAGCTGAGACC--GACGTGGATGA 905
 QY 273 CCTGGCTACTGTGGGAGAGCGGACAGCTCTCACTAGTCCGGGACAGCGTCAACCGGAGTA 332
 Db 906 CTTGG---CTGGGGGAGAGAGGAGAGCTCACTATCCGGGACAGCGGCTCAACCGGAGTA 962
 QY 333 CAGAACAGTTCAGAGAGCTTCCGAGAGAGATCAAAATCTGAGTGGCTCTCTCCCTCC 392
 Db 963 CAGAACAGTTCAGAGAGCTTCCGAGAGAGATCAAAATCTGAGTGGCTCTCTCCCTCC 1022
 QY 393 TTTCCCTGCCCCCGCCGACGCTGACAGAAAGCTGCTAACCCCTTACAAAGAG 452
 Db 1023 TTTCCCTGCCCCCGCCGACGCTGACAGAAAGCTGCTAACCCCTTACAAAGAG 1082
 QY 453 TCCAGACATCTAGCCAGGCTTCTAGCCCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 512
 Db 1083 TCCAGACATCTAGCCAGGCTTCTAGCCCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1142
 QY 513 CTCCAACTAGAGGCGCTTCACTCTATCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 572
 Db 1143 CTCCAACTAGAGGCGCTTCACTCTATCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1202
 QY 573 ATGATGCCAGATTTAGATTTGGCTCTGCTCACTTATCAGAGTACAGAGAGAGAGAGAG 632
 Db 1203 ATGATGCCAGATTTAGATTTGGCTCTGCTCACTTATCAGAGTACAGAGAGAGAGAGAG 1262
 QY 633 CAGGAAAGAAAGATCTTAAAGAACCACTGTGGGTCAAGGAGAAATGAGAGAGAGAGAG 692
 Db 1263 CAGGAAAGAAAGATCTTAAAGAACCACTGTGGGTCAAGGAGAAATGAGAGAGAGAGAG 1322
 QY 693 TGGCAAGCTCTGAG 752
 Db 1323 TGGCAAGCTCTGAG 1382
 QY 753 GGCAGAGGAGACAG 812
 Db 1383 GGCAGAGGAGACAG 1442
 QY 813 AGGTCAAG 872
 Db 1443 AGGTCAAG 1502
 QY 873 AAG 932
 Db 1503 AAG 1562
 QY 933 GAG 992
 Db 1563 GAG 1622
 QY 993 GCCAATCTATATATAGATGGGTTTTTCCATATACAGCTGGTGTGTAATAAATGACATGA 1052
 Db 1623 GCCAATCTATATATAGATGGGTTTTTCCATATACAGCTGGTGTGTAATAAATGACATGA 1682
 QY 1053 AACTCTGCGCTCTGCGGCTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1112
 Db 1683 AACTCTGCGCTCTGCGGCTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1742

QY 1113 GGGCTGTCCTTCTCCCTCCACAGGCGTGTGTTGTTGGGGCTGCTCCATGACAGAGAG 1172
 Db 1743 GGGCTGTCCTTCTCCCTCCACAGGCGTGTGTTGTTGGGGCTGCTCCATGACAGAGAG 1802
 QY 1173 ATCACTAACAGAGATGAGAACCCAGAGGATGATGGGCTTTGGGTCTCGAGGTTGAGC 1232
 Db 1803 ATCACTAACAGAGATGAGAACCCAGAGGATGATGGGCTTTGGGTCTCGAGGTTGAGC 1862
 QY 1233 CCCAGCTTCTTCCAGCTTCCCTCTCC--GGAGTACAGCTCTCCATCAATCCCTCTTTAA 1291
 Db 1863 CCCAGCTTCTTCCAGCTTCCCTCTCC--GGAGTACAGCTCTCCATCAATCCCTCTTTAA 1922
 QY 1292 TCTATGAATCTATAGAGCTCGGTGTGTGTAACA-ACACACCCCTATGTTGCTTCAAAAT 1350
 Db 1923 TCTATGAATCTATAGAGCTCGGTGTGTGTAACAACAACAACCCCTATGTTGCTTCAAAAT 1982
 QY 1351 ACTCAGCATTAACATGTTGTTAGGCAAAATTCAGAGCTTTCTCAATCAGATTTACAATC 1410
 Db 1983 ACTCAGCATTAACATGTTGTTAGGCAAAATTCAGAGCTTTCTCAATCAGATTTACAATC 2042
 QY 1411 TCCATTTTCATTAACGGGGAAACATCCCGAGGCACTGAGTGTGCTTGTGCACTGAA 1470
 Db 2043 TCCATTTTCATTAACGGGGAAACATCCCGAGGCACTGAGTGTGCTTGTGCACTGAA 2102
 QY 1471 GGTATGATCTGAACCCAGGGGTGTCAACNGTGTCTCAACTGCCCACTCTGGGCACTGA 1530
 Db 2103 GGTATGATCTGAACCCAGGGGTGTCAACNGTGTCTCAACTGCCCACTCTGGGCACTGA 2162
 QY 1531 GGAATATTTCCCTCATTTCTACCTCTCTAAGGCTATGAGCCCTCCAGGCTTTCAGC 1590
 Db 2163 GGAATATTTCCCTCATTTCTAAGGCTATGAGCCCTCCAGGCTTTCAGC 2222
 QY 1591 TGGGGGATGGGGGAGTATAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1650
 Db 2223 TGGGGGATGGGGGAGTATAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2282
 QY 1651 ATCAGCTTAAACCTCTGGGAAATGCTGTGCTGCCCCAGAGTACTTGTGTTGCTCTCCAG 1710
 Db 2283 ATCAGCTTAAACCTCTGGGAAATGCTGTGCTGCCCCAGAGTACTTGTGTTGCTCTCCAG 2342
 QY 1711 ATACAGAAAG 1770
 Db 2343 ATACAGAAAG 2402
 QY 1771 TCAGCTCTCTCCCTCATAGCTCTCTGTGTAAGTGTGCTTAAAGGTGCTCTCACTCCCA 1830
 Db 2403 TCAGCTCTCTCCCTCATAGCTCTCTGTGTAAGTGTGCTTAAAGGTGCTCTCACTCCCA 2462
 QY 1831 CGACTCTGGG--CCCTTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1889
 Db 2463 CGACTCTGGG--CCCTTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2522
 QY 1890 AACCTGTATACAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1923
 Db 2523 AACCTGTATACAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2556

RESULT 6
 AK095911 2142 bp mRNA linear PRI 30-JAN-2004
 LOCUS AK095911
 DEFINITION to Homo sapiens cDNA FLJ38592 fis, clone HEMR1000185, highly similar to Homo sapiens mRNA for cardiovascular heat shock protein.
 ACCESSION AK095911
 VERSION AK095911.1 GI:21755262
 KEYWORDS Oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Obase, M., Nishi, T., Shibahara, T., Tanaka, T.,

Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Magatsuma, M., Shitatori, A., Sudo, H., Hosoi, T., Kaku, Y., Kodaira, H., Kondo, H., Suganara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K., Nariwaka, M., Yamazaki, M., Nishimura, K., Ishibashi, T., Yamashita, H., Hirakawa, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yoshida, M., Hotta, T., Kusano, J., Kanehori, K., Takahashi-Fujii, A., Hara, Y., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R., Takeuchi, K., Arita, M., Imose, N., Mutsaers, K., Yuki, F., Hara, A., Sasaki, N., Aotaka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohara, N., Sano, S., Mori, S., Nomiyama, H., Satoh, N., Takami, S., Terahashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mitoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujimura, T., Ono, T., Yamada, K., Fujii, Y., Oaki, K., Hiroo, M., Ohmori, Y., Kawabata, A., Hiki, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Ohtani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Tada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S. Complete sequencing and characterization of 21,243 full-length human cDNAs Nat. Genet. 36 (1), 40-45 (2004)

14702039

2

Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hiroo, M., Shimizu, F., Wakebe, H., Ono, T., Hishigaki, H., Watanabe, T., Oaki, K., Sugiyama, T., Irie, K., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai, H., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Magatsuma, M., Murakami, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T. NEDO human cDNA sequencing project Unpublished

3 (bases 1 to 2142)

Isogai, T. and Yamamoto, J. Direct Submission Submitted (04-JUL-2002) Takao Isogai, FLI Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3973, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES

source

1. 2142

/organism="Homo sapiens"

/mol_type="cDNA"

/db_xref="taxon:9606"

/clone="HEART100185"

/issue_type="heart"

/clone_lib="HEART1"

/note="cloning vector: pME18SFL3"

ORIGIN

Query Match 83.6%; Score 1661; DB 9; Length 2142;

Best Local Similarity 95.6%; Pred. No. 0;

Matches 1809; Conservative 0; Mismatches 62; Indels 22; Gaps 9;

QY 33 GCCATCCCAACAAAGCCCGGTGGGCAAGCAATCAAGACCTTAGAGAGCGCTATGA 92

Db 269 GGCCTTCCAGCCGCCCGGTGGGCAAGCAATCAAGACCTTAGAGAGCGCTATGA 328

QY 93 GTTGCGGTGAGCGTGAAGACCTTCACTGAAGACATATGTGACCACTCCCAACA 152

Db 329 GTTGCGGTGAGCGTGAAGACCTTCACTGAAGACATATGTGACCACTCCCAACA 385

QY 153 AACCAACATGAGGTGGCGGCTGAGAAAGCTGGCGCTGAACGACACTGTACATGAA 212

Db 386 CAACCAATGAGGTGGCGGCTGAGAAAGCTGGCGCTGAACGACACTGTACATGAA 443

QY 213 CACCTTACGATACACAAAGTCCGAGATGGCGGAGAGCGTGAAGCCGAGCGTGGTA 272

Db 444 TTGCT 492

QY 273 CCTGCGTACTGTGCGGAGAGCGAGCGCTCACTGTCGCGGAGCGCTCACTGAGCA 332

Db 493 CCTGCGGAG 549

QY 333 CAGAACAGCTTCAGCAACCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 382

Db 550 CAGAACAGCTTCAGCAACCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 609

QY 393 TTTCCTGTCCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 452

Db 610 TTTCCTGTCCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 669

QY 453 TCCAGACATCTGAGCCGAGGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 512

Db 670 TCCAGACATCTGAGCCGAGGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 729

QY 513 CTCCCAACATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 572

Db 730 CTCCCAACATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 789

QY 573 ATGATGCGGAGATTTCAAGATTTGAGCTTCAAGATTTGAGCTTCAAGATTTGAG 632

Db 790 ATGATGCGGAGATTTCAAGATTTGAGCTTCAAGATTTGAGCTTCAAGATTTGAG 849

QY 633 CAGGAGAGAGATCTTAAGAACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 692

Db 850 CAGGAGAGAGATCTTAAGAACCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 909

QY 693 TGGGCAAGCTCTGAG 752

Db 910 TGGGCAAGCTCTGAG 969

QY 753 GGCAGAGGAG 812

Db 970 GGCAGAGGAG 1029

QY 813 AGCTCAAGGAG 872

Db 1030 AGCTCAAGGAG 1089

QY 873 AAG 932

Db 1090 AAG 1149

QY 933 GAGCAGAGCTGAG 992

Db 1150 GAGCAGAGCTGAG 1209

QY 993 GGCACATCTGATATGATGAGGATTTTTCATATACAGCTGATGATGATGATGATG 1052

Db 1210 GGCACATCTGATATGATGAGGATTTTTCATATACAGCTGATGATGATGATGATG 1269

QY 1053 AACTCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 1112

Db 1270 AACTCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 1329

QY 1113 GGGCTGCT 1172

Db 1330 GGGCTGCT 1389

QY 1173 ATGACCTAAGAGAGTGAAGGACGAGGATGATGGGGCTTTGGTGGCTTCGAGTTGGAC 1232
 DB 1390 ATGACCTAAGAGAGTGAAGGACGAGGATGATGGGGCTTTGGTGGCTTCGAGTTGGAC 1449
 QY 1233 CCCAGCTTCTTGGACCTTCCCTCC-GGAGTCACTTCACATCCCTCTTTAA 1291
 DB 1450 CCCAGCTTCTTGGACCTTCCCTCCGGGAGTCACTTCACATCCCTCTTTAA 1509
 QY 1292 TCTATGATCTATGAGTGGTGGTGTGTATCA-ACAGACCTTCATGCTTCTCTTAAT 1350
 DB 1510 TCTATGATCTATGAGTGGTGGTGTGTATCA-ACAGACCTTCATGCTTCTCTTAAT 1569
 QY 1351 ACTCAGATTACATGCTGTTGAGGACCAATTCAGAGCTTCTCAATCAGATTACATC 1410
 DB 1570 ACTCAGATTACATGCTGTTGAGGACCAATTCAGAGCTTCTCAATCAGATTACATC 1629
 QY 1411 TCCATTTTCATTAACGGGGAAACATCCCGAGCCACTGAGTGTGTGTCTCACTGAA 1470
 DB 1630 TCCATTTTCATTAACGGGGAAACATCCCGAGCCACTGAGTGTGTGTCTCACTGAA 1689
 QY 1471 GGTAGATCTGAACCCGAGGAGTCAACAGTGTGTCTCACTCCCACTCTGGCACTGA 1530
 DB 1690 GGTAGATCTGAACCCGAGGAGTCAACAGTGTGTCTCACTCCCACTCTGGCACTGA 1749
 QY 1531 GGAATATTTCCCTCATTTCTACCTCTCTAAGGCTATGCAACCCCTCCCACTCTTCAG 1590
 DB 1750 GGAATATTTCCCTCATTTCTACCTCTCTAAGGCTATGCAACCCCTCCCACTCTTCAG 1809
 QY 1591 TGGGGGATGGGGGAGTCAATAGAAAGCCCTCATCTCCATCTGGGATAGGACCTTCC 1650
 DB 1810 TGGGGGATGGGGGAGTCAATAGAAAGCCCTCATCTGGGATAGGACCTTCC 1869
 QY 1651 ATCAGCTTTACCTGGGAAATGCTGTGCTGCCCCAGTGACTGTGTGTCTCTCCAC 1710
 DB 1870 ATCAGCTTTACCTGGGAAATGCTGTGCTGCCCCAGTGACTGTGTGTCTCTCCAC 1929
 QY 1711 ATACAGAGAGAGGATGAGGAGGAGGATGAGGATGAGGATGAGGATGAGGATGAGG 1770
 DB 1930 ATACAGAGAGAGGATGAGGAGGAGGATGAGGATGAGGATGAGGATGAGGATGAGG 1989
 QY 1771 TCAAGCTTCTCCCTCCATGCTCTGTGAGTGTGCTTAAGGATGGCTCTCACTCCCA 1830
 DB 1990 TCAAGCTTCTCCCTCCATGCTCTGTGAGTGTGCTTAAGGATGGCTCTCACTCCCA 2049
 QY 1831 CCACTCTGGG-CCCTTGGGGAGAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1889
 DB 2050 CCACTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2109
 QY 1890 AACCTGTATACATAAAGAGAGCTTCACAG 1922
 DB 2110 AACCTGTATACATAAAGAGAGCTTCACAG 2142

RESULT 7

AX747447 2410 bp mRNA linear PART 20-JUN-2003
 LOCUS AX747447
 DEFINITION Sequence 972 from Patent EP1308459.
 ACCESSION AX747447 GI:32131835
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Homo sapiens (human)

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Taneichi, K. I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahara, K. and Masuho, Y.
 Full-length cDNA sequences
 Patent: EP 1308459-A 972 07-MAY-2003;
 JOURNAL Helix Research Institute (JP) ; Research Association for Biotechnology (JP)

FEATURES
 source
 Location/Qualifiers
 1.2410
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"

Query Match 83.6%; Score 1660.6; DB 6; Length 2410;
 Best Local Similarity 95.6%; Pred. No. 0;
 Matches 1808; Conservative 0; Mismatches 61; Indels 22; Gaps 9;

ORIGIN

QY 33 GCCCATCCCAACCAAGCCCGGTGGGAGGAGACATCAAGACCTAGAGAGAGCCTATGA 92
 DB 539 GGCCTTCCACGCGCGCCCGGTGGGAGGAGACATCAAGACCTAGAGAGAGCCTATGA 598
 QY 93 GTTGGGGTGAAGTGAAGAGCTTCACTGAAAGACATATTGTCACTCACTCAAGCA 152
 DB 599 GTTGGGGTGAAGTGAAGAGCTTCACTGAAAGACATATTGTCACTCACTCAAGCA 655
 QY 153 AACCAACATCGAGGTGGGAGGAGTGAAGAGTGGGAGGAGTGAAGAGTGAAGAG 212
 DB 656 CAACCAACATCGAGGTGGGAGGAGTGAAGAGTGGGAGGAGTGAAGAGTGAAGAG 713
 QY 213 CACCTTAAGACTTACCAAGTGGAGACTGCGGAGAGAGTGAAGAGTGAAGAGTGA 272
 DB 714 TTGGCT-----CAAGAGTGCAG-CTGCGGAGAGAGTGAAGAGTGAAGAGTGA 762
 QY 273 CCGGAGTCACTGAGGAGAGAGAGAGAGAGTCACTAGTCCGGGAGAGAGTGAAGAG 332
 DB 763 CTTGG---CTTGGGAGAGAGAGAGAGAGTCACTAGTCCGGGAGAGAGTGAAGAG 819
 QY 333 CAGAAACAGTCCAGAGAGAGTTCGAGAGAGAGATCAAAATGAGTGGCTCTCCCTTCC 392
 DB 820 CAGAAACAGTCCAGAGAGAGTTCGAGAGAGAGATCAAAATGAGTGGCTCTCCCTTCC 879
 QY 393 TTTCCCTGTCCCGCCCGCCAGAGAGTTCGAGAGAGAGTTCGAGAGAGAGTTCGAG 452
 DB 880 TTTCCCTGTCCCGCCCGCCAGAGAGTTCGAGAGAGAGTTCGAGAGAGAGTTCGAG 939
 QY 453 TCCAGAGAGTCCAGAGAGAGTTCGAGAGAGAGTTCGAGAGAGAGTTCGAGAGAG 512
 DB 940 TCCAGAGAGTCCAGAGAGAGTTCGAGAGAGAGTTCGAGAGAGAGTTCGAGAGAG 999
 QY 513 CTCCTTAAGAGAGAGTTCGAGAGAGTTCGAGAGAGAGTTCGAGAGAGAGTTCGAG 572
 DB 1000 CTCCTTAAGAGAGAGTTCGAGAGAGAGTTCGAGAGAGAGTTCGAGAGAGAGTTCGAG 1059
 QY 573 ATGATGCCAGATTTCAAGTTGGCTCGTCACTTAATCCAGAGTGAAGAGAGTTCGAG 632
 DB 1060 ATGATGCCAGATTTCAAGTTGGCTCGTCACTTAATCCAGAGTGAAGAGAGTTCGAG 1119
 QY 633 CAGGAGAGAGATTTAAAGAACCCAGTGTGGTCAAGGAGTTCGAGAGAGTTCGAG 692
 DB 1120 CAGGAGAGAGATTTAAAGAACCCAGTGTGGTCAAGGAGTTCGAGAGAGTTCGAG 1179
 QY 693 TGGGAGAGTTCGAGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAG 752
 DB 1180 TGGGAGAGTTCGAGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAG 1239
 QY 753 GGCAGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAG 812
 DB 1240 GGCAGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAG 1299
 QY 813 AGGTCAAGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAG 872
 DB 1300 AGGTCAAGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAG 1359
 QY 873 AAGAGAGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAG 932
 DB 1360 AAGAGAGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAG 1419
 QY 933 GAGCAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAGAGAGTTCGAG 992

CDS

```
/clone_11b="NTONG2"  
/note="cloning vector: pME18SFL3"  
126..863  
/note="unnamed protein product"  
/codon_start=1  
/protein_id="BAC03846.1"  
/db_xref="GI:21750828"  
/translation="MGRQVNPFRPGNCGOLDRESGTQQRVTAKHTWSHCQVLEMDXDS  
QCMGLDPMVAHVRSRQGLPGGGGIGLWQGGGPIATWBOGAPML  
RMASTLAQGDPEWEXALSMEDPFGSPMPSPLAPRPPGAGKTLGDAYEFA  
VDVDESPEDIVTSTNNHTEVAREKLAAGTYMNTFAHNCQLEPVDVPSVTSALRE  
DGLITRARRHPHEHVQTFREIKI"
```

ORIGIN

```
Query Match      83.6%; Score 1660.6; DB 9; Length 2410;  
Best Local Similarity 95.6%; Pred. No. 0;  
Matches 1808; Conservative 0; Mismatches 61; Indels 22; Gaps 9;  
  
QY 33 GCCCATCCCAACCAAGCCCGGTGGGAGGCAACATCAAGAAGCTTAGAGAGCGCTATGA 92  
DB 539 GGCCTTCCCAAGCCCGCGGTGGGAGGCAACATCAAGAAGCTTAGAGAGCGCTATGA 598  
QY 93 GTTGGCGGTGAGCTGAGAGACTCTCACTGAAGACATGATGTGTACCACTTCAAGACA 152  
DB 599 GTTGGCGGTGAGCTGAGAGACTCTCACTGAAGACATGATGTGTACCACTTCAAGACA 655  
QY 153 AACCAACATCGAGGTGGGGGGCTGAGAAAGCTGGGGCTGAACGAGACATGTAATGAAA 212  
DB 656 CAACCAACATCGAGGTGGGGGGCTGAGAAAGCTGGGGCTGAACGAGACATGTAATGAAA 713  
QY 213 CACCTTAGACATAACAAGTCCAGACTGCGGAGAGAGCTGAGACCGCGAGCGTGGCTGA 272  
DB 714 TTTCCT-----CACAAAGTCCAG-CTGCGGAGAGAGTGGAGCC--GACGTGGTGA 762  
QY 273 CCTGGCTACTGTGGGAGAGAGAGAGAGCTCACTAGTCCGGGAGCGCTACCCGGATA 332  
DB 763 CCGTGG--CTGTGGGAGAGAGAGAGAGAGCTCACTAGTCCGGGAGCGCTACCCGGATA 819  
QY 333 CAGAAACGCTCCAGCAAGCTTCCGAGAGAGATCAAAATCTGAGTCCCTCTCCCTTCCC 392  
DB 820 CAGAAACGCTCCAGCAAGCTTCCGAGAGAGATCAAAATCTGAGTCCCTCTCCCTTCCC 879  
QY 393 TTTCCTGATCCCGCGGAGAGAGAGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 452  
DB 880 TTTCCTGATCCCGCGGAGAGAGAGAGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 939  
QY 453 TCCAGAGATCTCAAGCCAGGTTTACGCCCCCAGCAGACCCCAAGACCCAGGTGAGCATC 512  
DB 940 TCCAGAGATCTCAAGCCAGGTTTACGCCCCCAGCAGACCCCAAGGTGAGCATC 999  
QY 513 CTCCCAACTAGAGGCGCTCACTATCAAGGAGAGAGAGAGAGCTCCCTGAGCTGAGCAC 572  
DB 1000 CTCCCAACTAGAGGCGCTCACTATCAAGGAGAGAGAGAGAGCTCCCTGAGCTGAGCAC 1059  
QY 573 ATGATGCCAATTTAGATTTGAGCTTCCGCTCACTTATCAAGTACAGAGGAGCTGAGGCT 632  
DB 1060 ATGATGCCAATTTAGATTTGAGCTTCCGCTCACTTATCAAGTACAGAGGAGCTGAGGCT 1119  
QY 633 CAGGAGAGAGAGATCTAAAGAACCACTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 692  
DB 1120 CAGGAGAGAGAGATCTAAAGAACCACTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1179  
QY 693 TGAGCAAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 752  
DB 1180 TGAGCAAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1239  
QY 753 GGCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 812  
DB 1240 GGCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1299  
QY 813 AGGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 872  
DB 1300 AGGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1359
```

```
QY 873 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 932  
DB 1360 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1419  
QY 933 GAGCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 992  
DB 1420 GAGCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1479  
QY 993 GGCACATCTGATATAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1052  
DB 1480 GGCACATCTGATATAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1539  
QY 1053 AACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1112  
DB 1540 AACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1599  
QY 1113 GAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1172  
DB 1600 GAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1659  
QY 1173 ATCACTTAACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1232  
DB 1660 ATCACTTAACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1719  
QY 1233 CCGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1291  
DB 1720 CCGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1779  
QY 1292 TCTATGAATCTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1350  
DB 1780 TCTATGAATCTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1839  
QY 1351 ACTCAGCATTAACATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1410  
DB 1840 ACTCAGCATTAACATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1899  
QY 1411 TCCATTTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1470  
DB 1900 TCCATTTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1959  
QY 1471 GGTAGATCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1530  
DB 1960 GGTAGATCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2019  
QY 1531 GGAATATTTCCCTCATTTCTACTCTCTTAAGGCTATGACCCCTCCCAAGCTTCTCCAGC 1590  
DB 2020 GGAATATTTCCCTCATTTCTACTCTCTTAAGGCTATGACCCCTCCCAAGCTTCTCCAGC 2079  
QY 1591 TGGGAGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1650  
DB 2080 TGGGAGATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2139  
QY 1651 ATCAGCTTAACCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1710  
DB 2140 ATCAGCTTAACCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2199  
QY 1711 ATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1770  
DB 2200 ATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2259  
QY 1771 TCAGCTCTCTCCCTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 1830  
DB 2260 TCAGCTCTCTCTCCCTCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 2319  
QY 1831 CCACTCTGGG--CCCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1889  
DB 2320 CCACTCTGGG--CCCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2379  
QY 1890 AACCTGTATACAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920  
DB 2380 AACCTGTATACAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2410
```

RESULT 9
LOCUS CQ783843 2701 bp DNA linear PAT 17-MAR-2004
DEFINITION Sequence 3983 from Patent EP139543.
ACCESSION CQ783843
VERSION CQ783843.1 GI:45503747
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Oka, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
Makamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
Koga, H.
TITLES Primers for synthesizing full length cDNA clones and their use
JOURNAL Patent: EP 139543-A 3983 10-MAR-2004;
Research Association for Biotechnology (JP)
FEATURES
source location/Qualifiers
1..2701
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
792..1151
/note="unnamed protein product"
/codon_start=1
/protein_id="CA86888.1"
/db_xref="GI:45503748"
/translation="MOGLHSLTAAHPAPRGAGNITKLDAYEPAVDYDPSPEP
LIVTISNHIWRAEKLADGTVMNTFAHKQLPBDVPTSVTSALREDSGLTIRAR
HPHTEHVQTFREIKI"
ORIGIN
Query Match 83.5%; Score 1658.8; DB 6; Length 2701;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1806; Conservative 0; Mismatches 59; Indels 23; Gaps 9;
QY 39 CCCAACCAAGCCCGGCGGAGGAGCAATCAAGACCTTAGAGAGCGCTATAGTTGC 98
DB 833 CACAGCCCGCCCGGCGGAGGAGCAATCAAGACCTTAGAGAGCGCTATAGTTGC 892
QY 99 GGTGACGTTGAGAGATTTCTCACTGAAACATCATTTGACCACTCCAAACCA 158
DB 893 GGTGACGTTGAGAGATTTCTCACTGAAACATCATTTGACCACTCCAAACCA 949
QY 159 CATGAGTGGGCGGCTGAGAAAGCTGCGGCTGAACGACACTTACATGAACACTT 218
DB 950 CATGAGTGGGCGGCTGAGAAAGCTGCGGCTGAACGACACTTACATGAACACTT 1007
QY 219 ACGACTACACAAAGTCCCAACTGCGGAGAGGTGACCCCGAGCGTGGTACCTGG 278
DB 1008 -----CAAGATGCCAG-CTGCCGAGAGAGGTGACCC--GACGTGGTGAACCTGG 1056
QY 279 CTACTGTGGGAG 338
DB 1057 ----CTTGGGAG 1113
QY 339 AGCTTCAG 398
DB 1114 AGCTTCAG 1173
QY 399 TGTCCCG 458
DB 1174 TGTCCCG 1233
QY 459 ACATTCAG 518
DB 1234 ACATTCAG 1293
QY 519 AACTAG 578
DB 1294 AACTAG 1353

QY 579 CCNAGATTTCAGATTGGCTCCGTCATTAATCCAGAGTACAGGGGCTGGGCTCAGGGA 638
DB 1354 CCNAGATTTCAGATTGGCTCCGTCATTAATCCAGAGTACAGGGGCTGGGCTCAGGGA 1413
QY 639 AGAAGATCTTAAAGAACCCACTGTGGGTGAGGGGAAATGGGACGAGAGACATATGGCA 698
DB 1414 AGAAGATCTTAAAGAACCCACTGTGGGTGAGGGGAAATGGGACGAGAGACATATGGCA 1473
QY 699 AGCTTCAG 758
DB 1474 AGCTTCAG 1533
QY 759 GGGACCGAG 818
DB 1534 GGGACCGAG 1593
QY 819 CAGGCAAGGGTGCCTATCTAAGTGAATTAATTCAGAGAGAGAGAGAGAGAGAGAG 878
DB 1594 CAGGCAAGGGTGCCTATCTAAGTGAATTAATTCAGAGAGAGAGAGAGAGAGAGAG 1653
QY 879 AGACAGCCGTGACGGTAACTTCCCTCTTACAGAGCTTCCAGAGCCCAAGCCAGAGAG 938
DB 1654 AGACAGCCGTGACGGTAACTTCCCTCTTACAGAGCTTCCAGAGCCCAAGCCAGAGAG 1713
QY 939 GCTCCGAG 998
DB 1714 GCTCCGAG 1773
QY 999 TCTGTATATATAGATGGGTTTTCATATACAGCTGTTGTTGAAAACTGATGAATCC 1058
DB 1774 TCTGTATATATAGATGGGTTTTCATATACAGCTGTTGTTGAAAACTGATGAATCC 1833
QY 1059 TGCGGCTGAG 1118
DB 1834 TGCGGCTGAG 1893
QY 1119 GTCCCTTCCTCCCAAGAGCTGTGTTCTGAGAGAGAGAGAGAGAGAGAGAGAG 1178
DB 1894 GTCCCTTCCTCCCAAGAGCTGTGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1953
QY 1179 TAAAGAGATGAAAG 1238
DB 1954 TAAAGAGATGAAAG 2013
QY 1239 TTCTTGACACTTCCCTCC-GGAGATCAGCTTCCATCATCCCTCTTTAATCTATG 1297
DB 2014 TTCTTGACACTTCCCTCC-GGAGATCAGCTTCCATCATCCCTCTTTAATCTATG 2073
QY 1298 AATCTATAGGCTGGGTGTGTAA-ACACAGCCCTATGTTGTCTTCAATATCTAG 1356
DB 2074 AATCTATAGGCTGGGTGTGTAA-ACACAGCCCTATGTTGTCTTCAATATCTAG 2133
QY 1357 CATTACATTTGTTAG 1416
DB 2134 CATTACATTTGTTAG 2193
QY 1417 TTCAATTAAGGGGAAATCCTCCGAGCACTGAGTGTGCTTTGTCACTGAAGTTAG 1476
DB 2194 TTCAATTAAGGGGAAATCCTCCGAGCACTGAGTGTGCTTTGTCACTGAAGTTAG 2253
QY 1477 ATCTGAAGCCAGGGGTGCAACGCTGCTCAACTCCCAAGCTTGGAGAGAGAGAGAG 1536
DB 2254 ATCTGAAGCCAGGGGTGCAACGCTGCTCAACTCCCAAGCTTGGAGAGAGAGAGAG 2313
QY 1537 TTTCCTCTCATTTAATCTCTTAAGGCTATGACCCCTCCAGAGCTTTCCAGCTGGGG 1596
DB 2314 TTTCCTCTCATTTAATCTCTCTTAAGGCTATGACCCCTCCAGAGCTTTCCAGCTGGGG 2373
QY 1597 ATGGGGAGAGTCAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1656
DB 2374 ATGGGGAGAGTCAATAG 2433

Accession	Sequence	Position
QY	CTTAAACCTGGGAATGCTCTGCGCCCACTACTCTTGTTTCGTTCCCAATACAG	1716
Db	CTTAAACCTGGGAATGCTCTGCGCCCACTACTCTTGTTTCGTTCCCAATACAG	2493
QY	1717 AAGCAGGGTGGAGGGGATGAGTCTCAGTTAGCAGGGATCCCCAAGGCAATGACCC	1776
Db	2494 AAGCAGGGTGGAGGGGATGAGTCTCAGTTAGCAGGGATCCCCAAGGCAATGACCC	2553
QY	1777 TCTCTCCCTCCATGCTCTCTGATGAGTATGCTTAGGGTGGGCTCTCACTCCCACTC	1836
Db	2554 TCTCTCCCTCCATGCTCTCTGATGAGTATGCTTAGGGTGGGCTCTCACTCCCACTC	2613
QY	1837 TGGGCCC-TTGGGGGAGGACTGGGGAGGGGGCCGTGGAGAGCCCTGACGCTGAACT	1894
Db	2614 TGGGCCCCCTTGGGGGAGGAGCTGGGGAGGGGGCCGTGGAGAGCCCTGACGCTGAACT	2673
QY	1895 GTATACACAATAAAGAGACAGTCTCAACAG	1922
Db	2674 GTATACACAATAAAGAGACAGTCTCAACAG	2701

RESULT	10
LOCUS	BD127861
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.
ACCESSION	BD127861
VERSION	BD127861.1 GI:23222806
KEYWORDS	JP 2002017375-A/3292.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE 1 (bases 1 to 2701)
AUTHORS Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002017315-A 3282 22-JAN-2002;

COMMENT

EN UP 20020173-5-4,3292
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC
 C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
 10,
 PC, C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof PH Key
 Location/Qualifiers
 FT CDS (792)..(1148).

```
FEATURES
  source
    location/Qualifiers
      1..2701
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"

ORIGIN
```

Query Match	83.5%	Score 1658.8	DB 6	Length 2701
Best Local Similarity	95.7%	Pred. No. 0		
Matches 1806	Conservative 0	Mismatches 59	Indels 23	Gaps 9

Qy	39	CCCAACCAAGCCCCCGGTGGGCAAGGACATCAAGACCTTAGAGAGACGCTTAGATTGG	98
Db	833	CACAGCCCCCGGTGGGAGGAGGACATCAAGACCTTAGAGAGACGCTTAGATTGG	892
Qy	99	GGTGACGTGAGAGACTTCTCACTGAAGACATCTTGTCAACCTTCAACCAACCA	158
Db	893	GGTGACGTGAGAGACTTCTCACTGAAGACATCTTGTCAACCTTCAACCAACCA	949

OY	159	TATGAGATGGGAGGCTGAGAAACCTGAGGCTGAAAGGACACTGATCATGAACAACCTT	218
Db	950	CATGAGATGGGG--CTGAAAGCTGGCGGTACCGCAGCCGTATGAAACACTTGGCT	1007
OY	219	ACGACTACACAGTGCACAGACTGCGGAGAGCTGAGCCCGAGCGTGGTACCTTGG	278
Db	1008	-----CACAAAGTGCAG--CTGCGGAGAGCGTGGAGCC--AGCTGGGTGACCTTGG	1056
OY	279	CTACTGTGGGAGGACCGAGACGCTCATCTAGTCCGAGGACCGGTACCCCGATACAGAAC	338
Db	1057	---CTGTGGGAGGAGAGCGAGCTTCACTATCCGGGACAGGCGTACCCCGATACAGAAC	1113
OY	339	ACGTCCAGCAGACCTTCCGAGCGGAGATCAAAATCTGAGTGCCTTCCCTTCC	398
Db	1114	ACGTCCAGCAGACCTTCCGAGCGGAGATCAAAATCTGAGTGCCTTCCCTTCC	1173
OY	399	TGTCCCCCGCCCGACGCTGCGACGAGAAAGCTGCTAAACCCATTAACAAGCTCAGAG	458
Db	1174	TGTCCCCCGCCCGACGCTGCGACGAGAAAGCTGCTTAACCCATTAACAAGCTCAGAG	1233
OY	459	ACATCTCAGCCAGGATTTAGCCCCCAAGCAGCCAGACCCAGGTGACCATCTTCCA	518
Db	1234	ACATCTCAGCCAGGATTTAGCCCCCAAGCAGCCAGACCCAGGTGACCATCTTCCA	1293
OY	519	AACTAGGAGCCCTCCACTCTATCCAGGGGACGGGAGATCCCGTCCCTGACACTGATG	578
Db	1294	AACTAGGAGCCCTCCACTCTATCCAGGGGACGGGAGATCCCTGCTGAGACATGATG	1355
OY	579	CCGAGATTTGAGATTTGGCCCTCCGCTCACTTATCCAGAGTACAGGGGCTGGGGTACAGGA	638
Db	1354	CCGAGATTTGAGATTTGGCCCTCCGCTCACTTATCCAGAGTACAGGGGCTGGGGTACAGGA	1413
OY	639	AGGAAGATCTTAAAGAACCACTGTGGGTCAAGGGGATGGGACACAGCAGAGACATATGGGCA	698
Db	1414	AGGAAGATCTTAAAGAACCACTGTGGGTCAAGGGGATGGGACACAGCAGAGACATATGGGCA	1477
OY	699	AGCTCTGAGAGACAGACAGACAGACAAACCTCTGATCTATGAAGTCTCTCAGGGCAG	758
Db	1474	AGCTCTGAGAGACAGACAGACAGACAAACCTCTGATCTATGAAGTCTCTCAGGGCAG	1533
OY	759	GGGACACAGGACCTTGGAACTCTTTGGCCAAAGGGGATGGGAGAGACAGAGGAAGTCA	818
Db	1534	GGGACACAGGACCTTGGAACTCTTTGGCCAAAGGGGATGGGAGAGAGACAGAGGAAGTCA	1593
OY	819	CAGACAAAGGGGCTACTATAGTGAAGTAATTCGCCAGGGGCTCAGCAAGGCCAAGAG	878
Db	1594	CAGACAAAGGGGCTACTATAGTGAAGTAATTCGCCAGGGGCTCAGCAAGGCCAAGAG	1655
OY	879	AGACAGCCGTGACGGTAACTTCCCTCTACAGCTTCAAGCCACGCCAGCGAGCAG	938
Db	1654	AGACAGCCGTGACGGTAACTTCCCTCTACAGCTTCAAGCCACGCCAGCGAGCAG	1713
OY	939	GCTGCTGCGCCACCCCGTGGCCCGACGAGCTGGGTGCGCAGGGCAGCATCCACA	998
Db	1714	GCTGCTGCGCCACCCCGTGGCCCGACGAGCTGGGTGCGCAGGGCAGCATCCACA	1777
OY	999	TCTGTATATAGATGGGTTTTTCCAAATACAGCTGATGTGTAAGAACTGCATGAATCTC	1057
Db	1774	TCTGTATATAGATGGGTTTTTCCAAATACAGCTGATGTGTAAGAACTGCATGAATCTC	1833
OY	1059	TGCGGCTCGGAGCTGCTGGGGGCTCCAGAGGACAAAGGCAAGTGGGTTGGGGTGGGGCTG	1118
Db	1834	TGCGGCTCGGAGCTGCTGGGGGCTCCAGAGGACAAAGGCAAGTGGGTTGGGGTGGGGCTG	1893
OY	1119	GTCCTTCTCCTCCCAAGGCTGTGTTGTTGAGGGCTGCTCCCATGCAAGCATCATCC	1178
Db	1894	GTCCTTCTCCTCCCAAGGCTGTGTTGTTGAGGGCTGCTCCCATGCAAGCATCATCC	1953
OY	1179	TAAACAGAGATGAAGCCAGGGCATGAGATGGGGCTTTGGGTCTCTGAGAGTTGAAACCCAGC	1238
Db	1954	TAAACAGAGATGAAGCCAGGGCATGAGATGGGGCTTTGGGTCTCTGAGAGTTGAAACCCAGC	2013

Db 1059 AGCTCCAGAGACCTTCCGGAGGAGATGAAATCTGAGTGCCTCCCTCCCTTCC 1118
Qy 399 TGTCCCCCGCCGACGCTGCGACGAAGCTCTGTAACCCCTTACACAGCTCCAG 458
Db 1119 TGTCCCCCGCCGACGCTGCGACGAAGCTCTGTAACCCCTTACACAGCTCCAG 1178
Qy 459 ACATCTCAGCCAGTCTTAGCCCCCAGCAGCCAGACCCAGAGTGAACATCTCCCA 518
Db 1179 ACATCTCAGCCAGTCTTAGCCCCCAGCAGCCAGACCCAGAGTGAACATCTCCCA 1238
Qy 519 AACTTAGGCCCCCTCAGCTTATCCAGGCGAGGCGAGGAGATCCCTGCTGACACATGATG 578
Db 1239 AACTTAGGCCCCCTCAGCTTATCCAGGCGAGGCGAGGAGATCCCTGCTGACACATGATG 1298
Qy 579 CCCAGATTCAGATTTGGCCTCCGTCACCTTAATCCAGATACAGGGGCTGGGTCAGGA 638
Db 1299 CCCAGATTCAGATTTGGCCTCCGTCACCTTAATCCAGATACAGGGGCTGGGTCAGGA 1358
Qy 639 AGGAAGATCTAAAGAACCACTGTGGGTACAGGGAAATGGAGCAAGACATATGGCA 698
Db 1359 AGGAAGATCTAAAGAACCACTGTGGGTACAGGGAAATGGAGCAAGACATATGGCA 1418
Qy 699 AGCTTCGAGGAGCAGACAGACAAACCTCTGATCTATGAGTCTTGAGGGGCAAG 758
Db 1419 AGCTTCGAGGAGCAGACAGACAAACCTCTGATCTATGAGTCTTGAGGGGCAAG 1478
Qy 759 GGGACCAAGGACCTGGAACCTTCTTGCCCAAGGGAGTGGAGAGACAGAGGAAAGTCA 818
Db 1479 GGGACCAAGGACCTGGAACCTTCTTGCCCAAGGGAGTGGAGAGACAGAGGAAAGTCA 1538
Qy 819 CAGGCAAGGAGTGTCTTCTTAAGTGAACCTTAATGGCCGAGGGCTACAGAAAGCCAAAGG 878
Db 1539 CAGGCAAGGAGTGTCTTCTTAAGTGAACCTTAATGGCCGAGGGCTACAGAAAGCCAAAGG 1598
Qy 879 AGACAGCCGTCAGCGTAACTTCCCTTACAGAGCTCCAAAGCCCAAGCCAGCAGAG 938
Db 1599 AGACAGCCGTCAGCGTAACTTCCCTTACAGAGCTCCAAAGCCCAAGCCAGCAGAG 1658
Qy 939 GCTGCTGCCCCAGCCCGTCCCGCCAGCGCTGTGTCAGGCGAAGACCATCCACA 998
Db 1659 GCTGCTGCCCCAGCCCGTCCCGCCAGCGCTGTGTCAGGCGAAGACCATCCACA 1718
Qy 999 TCTGTATAGATGGGGTTTTCCTCAATACAGCTGGTGGTGAATAATGCAATGAATCC 1058
Db 1719 TCTGTATAGATGGGGTTTTCCTCAATACAGCTGGTGGTGAATAATGCAATGAATCC 1778
Qy 1059 TGGCGTCTGCGCTGCTGCGGCTTCAGGCAAGGCCAAATGGGGTTGGGGTGGGCTG 1118
Db 1779 TGGCGTCTGCGCTGCTGCGGCTTCAGGCAAGGCCAAATGGGGTTGGGGTGGGCTG 1838
Qy 1119 GTCTCTTCCCTCCCAAGGCGCTGTGTTCTGTTGAGGCTGCTCCAGCAGACAGATCAC 1178
Db 1839 GTCTCTTCCCTCCCAAGGCGCTGTGTTCTGTTGAGGCTGCTCCAGCAGACAGATCAC 1898
Qy 1179 TAACAGAGATGAGGAGGAGGAGTGGGCTTGGGCTCTCGAGGTTGGAGCCCGAGC 1238
Db 1899 TAACAGAGATGAGGAGGAGGAGTGGGCTTGGGCTCTCGAGGTTGGAGCCCGAGC 1958
Qy 1239 TTCTTGCCCACTTCCCTCCGCGAGCTGCTCCATCCATCCCTCTTAAATCTATG 1297
Db 1959 TTCTTGCCCACTTCCCTCCGCGAGTGAAGCTGCTCCATCCATCCCTCTTAAATCTATG 2018
Qy 1299 AATCTAAGGCTCGGTGTGTATACA-ACACACCCCTAATGTTGCTTCAAAATCTAG 1356
Db 2019 AATCTAAGGCTCGGTGTGTATACA-ACACACCCCTAATGTTGCTTCAAAATCTAG 2078
Qy 1357 CATTAACATTTGTTGAGGCGAAATTCAGAGCTTTCGAAATCAGATTTACATCTCATT 1416
Db 2079 CATTAACATTTGTTGAGGCGAAATTCAGAGCTTTCGAAATCAGATTTACATCTCATT 2138
Qy 1417 TTCATTAAGGGGAAACATCCCGAGCACTAGTGTGCTTGTCACTGAAGGTTAG 1476

Db 2139 TTCATTAAGGGGAAACATCCCGAGCACTGAGTGTGCTTGTCACTGAAGGTTAG 2198
Qy 1477 ATCTGAACCCAGGGTGCACACNGTGTCTCACTCCCACTCTGTCAGCATGAGAGTA 1536
Db 2199 ATCTGAACCCAGGGTGCACACAGTGTCTCACTCCCACTCTGTCAGCATGAGAGTA 2258
Qy 1537 TTTCCTCATTTACTCTCTTAAGGCTATGCAACCCCTCCAGTCTTCCAGCTGGGG 1596
Db 2259 TTTCCTCATTTACTCTCTTAAGGCTATGCAACCCCTCCAGTCTTCCAGCTGGGG 2318
Qy 1597 ATGGGGGAGTCAATAGGAAAGGCCCATATCTCCCATCTGGGATAGGGACCTTCAATCAG 1656
Db 2319 ATGGGGGAGTCAATAGGAAAGGCCCATATCTCCCATCTGGGATAGGGACCTTCAATCAG 2378
Qy 1657 CTTAACCTTGGGAATCCCTGCTGCCCGCAGTACTCTTGTCTGCTCCCATATAG 1716
Db 2379 CTTAACCTTGGGAATCCCTGCTGCCCGCAGTACTCTTGTCTGCTCCCATATAG 2438
Qy 1717 AACGAGGTGAGGAGGAGGAGGAGTGTCTCAGTTAGAGAGGAGTCCAGAGGCAAGTACGC 1776
Db 2439 AACGAGGTGAGGAGGAGGAGGAGTGTGTCTCAGTTAGAGAGGAGTCCAGAGGCAAGTACGC 2498
Qy 1777 TCTTCCCTCATGCTCTGTGTCAGTGTGCTTGAAGGCTTGAAGGCTTCCATCCACATCAG 1836
Db 2499 TCTTCCCTCATGCTCTGTGTCAGTGTGCTTGAAGGCTTGAAGGCTTCCATCCACATCAG 2558
Qy 1837 TGGG-CCCTTGGGAGAGACTGTGGAGAGGGGCGGTGGAGAGGCTTGAAGGCTTGAAGGCTG 1895
Db 2559 TGGGCGCTTGGGAGAGACTGTGGAGAGGGGCGGTGGAGAGGCTTGAAGGCTTGAAGGCTG 2618
Qy 1896 TATACACATTAATAGGACAGTCTACAGA 1923
Db 2619 TATACACATTAATAGGACAGTCTACAAA 2646

RESULT 13
HSM806124 2163 bp mRNA linear PRI 17-JUN-2003
LOCUS
DEFINITION
Homo sapiens mRNA; cDNA DKFZp686A14192 (from clone DKFZp686A14192); complete cds.
ACCESSION
BX537990
VERSION
BX537990.1 GI:31874066
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2163)
Poustka, A., Albert, R., Moosmayer, P., Schupp, I., Wellenreuther, R., Mewes, H. W., Weill, B., Amid, C., Oesinger, A., Fobo, G., Han, M. and Wiemann, S.
AUTHORS
TITLE
Direct Submission
JOURNAL
Submitted (17-JUN-2003) MIPs, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
COMMENT
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de; sequenced by DKFZ (German Cancer Research Center, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp686A14192) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cdna/>.
FEATURES
source
1..2163
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="1p36.13"
/clone="DKFZp686A14192"
/issue_type="human esophagus tumor"
/clone_id="686 (synonym: hicc3). Vector pSport1_Sfi; host

gene DH10B; sites SfiIA + SfiIB
/dev stage="adult"
1..2163
/gene="DKFZp686A14192"
<1..577
/gene="DKFZp686A14192"
/note="Heat-shock protein, beta-7, differentially spliced"
/codon start=2
/product="hypothetical protein"
/protein_id="CAD97349.1"
/db_xref="GI:31874067"
/translation="ARPEACATLSCPAFARLRARRRRPWRHRSSTRPARSRHS
SSSSSSSSSSSARALAQDPPEKALSWFSDFGFPMSPHEPLAFPARPAGANIK
TLGDAYEFAVDYDFSPEDILVITSNHIELADGTVMNTFAHKCOLPEDVPTSVTS
ALRBDGLTIRRHHPTEHVQOTFRTEIKI"
2107..2112
/gene="DKFZp686A14192"
2152
/gene="DKFZp686A14192"

polYA_signal
polYA_site

ORIGIN

Query Match 82.9%; Score 1646.8; DB 9; Length 2163;
Best Local Similarity 94.9%; Pred. No. 0;
Matches 1799; Conservative 0; Mismatches 59; Indels 37; Gaps 8;

33 GCCCATCCCAACCAAGCCCGGTGGGCGAGCAACATCAGACCCCTAGAGAGACGCTATGA 92
268 GGCCTTCCAGCCCGCCCGGTGGGCGAGCAACATCAGACCCCTAGAGAGACGCTATGA 327
QY GTTTGGGTGGAGGTGAGACTTCTCACTGAAGACATCATTTGACCAACCTCCAAACA 152
Db GTTTGGGTGGAGGTGAGACTTCTCACTGAAGACATCATTTGACCAACCTCCAAACA 387
QY 328 GTTTGGGTGGAGGTGAGACTTCTCACTGAAGACATCATTTGACCAACCTCCAAACA 387
153 AACCAACATGAGAGGTGGGCGGTGAGAAAGCTGGCGCTGAACGACACTGATCAATGAA 212
Db CCACATC-----GAGCTGGCGGCTGACGGACCGTCATGAACAC 427
QY 213 CACCTTACGACTACACAAAGTCCAGAGTGGCGAGAGAGCTGGAGAGCCCGAGCGGTGA 272
428 TTGGCT-----CACAAATGCCAG-CTGCCGAGAGCGTGAACCC--GACCTGAGTA 476
QY 273 CCTGGCTACTGTGGGAGAGCGGACGCTCACTAGTCCGGGACGCGTCAACCCGAT 332
Db 477 CCTGG-----CTGGCGGAGAGAGCGGACGCTCACTAGTCCGGGACGCGTCAACCCGAT 533
QY 333 CAGAACGCTTCAGACACCTTCCGAGCGAGATCAAAATCTGAGTGCCTCTCCCTCC 392
534 CAGAACGCTTCAGACACCTTCCGAGCGAGATCAAAATCTGAGTGCCTCTCCCTCC 593
QY 393 TTTTCCGTGCCCCCGCCAGCGCTGCGAGCAAGCTGCTAACCCTTCAACACAG 452
594 TTTTCCGTGCCCCCGCCAGCGCTGCGAGCAAGCTGCTAACCCTTCAACACAG 653
Db 453 TCCAGACATCTCAGCCAGGTTCTAGCCCCCAGACGCCAGACCCCAAGTGAACATC 512
654 TCCAGACATCTCAGCCAGGTTCTAGCCCCCAGACGCCAGACCCCAAGTGAACATC 713
QY 513 CTCCCAACTAGGCGCTTCACTCTATCCAGGGAGGCGAGGACTCCCGGCTGACAC 572
714 CTCCCAACTAGGCGCTTCACTCTATCCAGGGAGGCGAGGACTCCCGGCTGACAC 773
Db 573 ATGATGCCAGATTTCAGATTGGCTCCGCTCACTTATCCAGAGTACAGGGGTGGGGT 833
QY 774 ATGATGCCAGATTTCAGATTGGCTCCGCTCACTTATCCAGAGTACAGGGGTGGGGT 833
Db 633 CAGGAGAGAGATCTTAAAGAACCCACTGTGGGTCAAGGAGATGGGACAGAGACAT 692
QY 634 CAGGAGAGAGATCTTAAAGAACCCACTGTGGGTCAAGGAGATGGGACAGAGACAT 893
834 CAGGAGAGAGATCTTAAAGAACCCACTGTGGGTCAAGGAGATGGGACAGAGACAT 893
QY 693 TGGCAAGCTCTGAGGACAGACAGACAAACCTCTGATATGAAGTCTTGAC 752
Db 894 TGGCAAGCTCTGAGGACAGACAGACAAACCTCTGATATGAAGTCTTGAC 953

QY 753 GGCAGAGGACACAGGACCTTGAACCTCTTGGCCAAAGGAGTGGAGACAGAGGA 812
Db 954 GGCAGAGGACACAGGACCTTGAACCTCTTGGCCAAAGGAGTGGAGACAGAGGA 1013
QY 813 AGGTACAGGACAGGAGTCCCTATCTAAGTGAACATAATGGCCGAGGGCTCAGAGGCC 872
Db 1014 AGGTACAGGACAGGAGTCCCTATCTAAGTGAACATAATGGCCGAGGGCTCAGAGGCC 1073
QY 873 AAGAGAGACAGCGGTGAGGTAACCTTCCCTTACAGGCTCCAAAGCCCAAGCCAGC 932
Db 1074 AAGAGAGACAGCGGTGAGGTAACCTTCCCTTACAGGCTCCAAAGCCCAAGCCAGC 1133
QY 933 GAGCAGGCTGCTGCCACCCCGTGGCCCGCCAGCCAGAGTGTGTGCCAGAGGACAT 992
Db 1134 GAGCAGGCTGCTGCCACCCCGTGGCCCGCCAGCCAGAGTGTGTGCCAGAGGACAT 1193
QY 993 GCCACATCTGATATAGATGGGGTTTTTCCATACAGCTGGTGTGTAATAACTGATGA 1052
Db 1194 GCCACATCTGATATAGATGGGGTTTTTCCATACAGCTGGTGTGATTAACATGATGA 1253
QY 1053 AACTCTGCGCTGCTGCGCTGCTGGGCTCTCAGGACAGGCAAGTGGGTTGGGGTG 1112
Db 1254 AACTCTGCGCTGCTGCGCTGCTGGGCTCTCAGGACAGGCAAGTGGGTTGGGGTG 1313
QY 1113 GGGCTGTGCTTCTCCCTCCAGAGGCTGTGTTCTGTTGGGGCTGTCCATGACAGACG 1172
Db 1314 GGGCTGTGCTTCTCCCTCCAGAGGCTGTGTTCTGTTGGGGCTGTCCATGACAGACG 1373
QY 1173 ATCACTTAAAGAGATGAGAACCCAGGAGATGATGGGCTTTGGGTTCTGAGGTTGAC 1232
Db 1374 ATCACTTAAAGAGATGAGAACCCAGGAGATGATGGGCTTTGGGTTCTGAGGTTGAC 1433
QY 1233 CCGAGCTTCTGCGACCTTCCCTTC-GGAGTACGCTCTCCATCAATCCCTCTTAA 1291
Db 1434 CCGAGCTTCTGCGACCTTCCCTTC-GGAGTACGCTCTCCATCAATCCCTCTTAA 1493
QY 1292 TCTATGAATCTATAGGCTCGGTGTGTGAACA-ACACACCCCTATGTTGCTTCAAT 1350
Db 1494 TCTATGAATCTATAGGCTCGGTGTGTGAACA-ACACACCCCTATGTTGCTTCAAT 1553
QY 1351 ACTGACATTAACATTTGTTGAGGCAATTCAGAGCTTTCTCAATCAGATTTCATC 1410
Db 1554 ACTGACATTAACATTTGTTGAGGCAATTCAGAGCTTTCTCAATCAGATTTCATC 1613
QY 1411 TCCATTTTCAATTAAGGGGAAACATCCCGAGCCCTGAGTGTGCTTGTGCTGAA 1470
Db 1614 TCCATTTTCAATTAAGGGGAAACATCCCGAGCCCTGAGTGTGCTTGTGCTGAA 1673
QY 1471 GGTAGATCTGAACCCAGGAGTGTCAACNGTGTCTCAACTCCCACTTGGGACATGA 1530
Db 1674 GGTAGATCTGAACCCAGGAGTGTCAACNGTGTCTCAACTCCCACTTGGGACATGA 1733
QY 1531 GAGATATTTCCCTCATTTCTACTCTCTTAAGGCTATGACCCCTCCCAAGCTTCCAGC 1590
Db 1734 GAGATATTTCCCTCATTTCTACTCTCTTAAGGCTATGACCCCTCCCAAGCTTCCAGC 1793
QY 1591 TGGGGGATGGGGGAGTATAGAGAAAGCCCATCTCCATCTGGGATAGGACCTTCC 1650
Db 1794 TGGGGGATGGGGGAGTATAGAGAAAGCCCATCTCCATCTGGGATAGGACCTTCC 1853
QY 1651 ATCAGCTTAAACCTTGGGAAATGCTGCTGCCCCAGTGAATCTTGGTTGGTCTCCAC 1710
Db 1854 ATCAGCTTAAACCTTGGGAAATGCTGCTGCCCCAGTGAATCTTGGTTGGTCTCCAC 1913
QY 1711 ATACAGAGCAGGAGTGAAGGAGGAGGAGTGTGAGTTAGAGAGGCTCCCAAGGCAAG 1770
Db 1914 ATACAGAGCAGGAGTGAAGGAGGAGGAGTGTGAGTTAGAGAGGCTCCCAAGGCAAG 1973
QY 1771 TGAAGCTCTCTCCCTCAATGCTCTTGTGCTGAGTGTGCTTGAAGTGTGCTCTCACTCCA 1830
Db 1974 TGAAGCTCTCTCCCTCAATGCTCTTGTGCTGAGTGTGCTTGAAGTGTGCTCTCACTCCA 2033
QY 1831 CCACTCTGGG-CCCTTGGGGAGAGCTGGGAGGAGGCGGTGGAGAGAGCCTGACGCTGG 1889

Db 2034 CCACTCTGGGCCCCCTTGGGGAGAGACTGGGAGGGGCGCTGGAGAGCCCTGACGCTGG 2093

Qy 1890 AACCTGTATACACATATAAGACAGTCTCAGAGA 1923

Db 2094 AACCTGTATACACATATAAGACAGTCTCAGAGA 2127

RESULT 14

AF155908 2152 bp mRNA linear PRI 27-DEC-1999

LOCUS AF155908

DEFINITION Homo sapiens cardiovascular heat shock protein mRNA, complete cds.

ACCESSION AF155908

VERSION AF155908.1 GI:6635998

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 2152)

REFERENCE Krief, S., Faivre, J.-F., Robert, P., Le Douarin, B., Brunet-Larignon, N., Leftere, I., Bouzyk, M.M., Anderson, K.M., Greller, L.D., Tobin, F.L., Souchet, M. and Brill, A. Identification and characterization of cvhsp. A novel human small stress protein selectively expressed in cardiovascular and insulin-sensitive tissues

TITLE JOURNAL Submitted (02-JUN-1999) Cardiovascular Pharmacology, Smithkline Beecham, 4 rue du Chesnay-Beauregard, Saint-Gregoire 35760, France

FEATURES

source Location/Qualifiers

1..2152

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/chromosome="1"

/map="1p36.23-p34.3 between markers DIS434 and DIS507"

75..587

/note="25 kDa heat stress protein; cvhsp; preferentially expressed in cardiovascular and insulin-sensitive tissues; interacts with C-terminal domain of alpha-filamin or actin-binding-280, Abp-280"

/codon_start=1

/product="cardiovascular heat shock protein"

/protein_id="AA020022.1"

/db_xref="GI:6635998"

/translation="MSHRTSSSTPFRARSPHSSSSSSSTSSASRALPADPMEKALSMFDDFGSPMRPSEPLAPAPAGAGNKGIDAYEFAVDYRDPSPEDIIYTTN NHIEVRAEKLIADGTVMNTPAHKQLPEDVDPSTVTSALREDSLTIRARHPTHEV QOTFTEYKI"

ORIGIN

polyA_signal 2115..2120

Query Match 82.7%; Score 1642.8; DB 9; Length 2152;

Best Local Similarity 95.6%; Pred. No. 0;

Matches 1811; Conservative 0; Mismatches 59; Indels 24; Gaps 11;

Qy 33 GCCCATCCCAACCAAGCCCGGTGGGACAGCAATCAAGACCTGAGAGACGGCTATGA 92

Db 263 GGCCTTCCAGCCCGCCCGGTGGGACAGCAATCAAGACCTGAGAGACGGCTATGA 322

Qy 93 GTTGGGAG 152

Db 323 GTTGGGAG 379

Qy 153 AACCAACATGAG 212

Db 380 CAACCAATGAG 437

Qy 213 CACCTTACGAGCTACACAG 272

Db 438 TTGCT-----CAG 486

Qy 273 CTTGGGCTACTGTGCGAG 332

Db 487 CTTGGG---CTTGGGAG 543

Qy 333 CAGAACAGCTCCAG 392

Db 544 CAGAACAGCTCCAG 603

Qy 393 TTTTCCTGTCCTCCCGCCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 452

Db 604 TTTTCCTGTCCTCCCGCCCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 663

Qy 453 TCCAGAGAGATTCAG 512

Db 664 TCCAGAGAGATTCAG 722

Qy 513 CTCCTCAAACTAG 572

Db 723 CTCCTCAAACTAG 782

Qy 573 ATGATGCCAGATTTAGATTTAGATTTAGATTTAGATTTAGATTTAGATTTAGATTTAG 632

Db 783 ATGATGCCAGATTTAGATTTAGATTTAGATTTAGATTTAGATTTAGATTTAGATTTAG 842

Qy 633 CAGGAG 692

Db 843 CAGGAG 902

Qy 693 TGGGCAAGCTCTCAG 752

Db 903 TGGGCAAGCTCTCAG 962

Qy 753 GCGAAG 812

Db 963 GCGAAG 1022

Qy 813 AGGTCAAG 872

Db 1023 AGGTCAAG 1082

Qy 873 AAG 932

Db 1083 AAG 1142

Qy 933 GAGCAG 992

Db 1143 GAGCAG 1202

Qy 993 GCGACATCTGTATATAGATGGGTTTTCATATACAGCTGTGTTGGTGAATACTGATGA 1052

Db 1203 GCGACATCTGTATATAGATGGGTTTTCATATACAGCTGTGTTGGTGAATACTGATGA 1262

Qy 1053 AACTCTGCGCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 1112

Db 1263 AACTCTGCGCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 1322

Qy 1113 GGGCTGTCT 1172

Db 1323 GGGCTGTCT 1382

Qy 1173 ATCACTTAACAGAGATGAG 1232

Db 1383 ATCACTTAACAGAGATGAG 1442

Qy 1233 CCGAGCTTTTGGCACTTCCCTCC-GGAGTCAAGCTCTCCATCCATCCCTCTTTTAA 1291

Db 1443 CCGAGCTTTTGGCACTTCCCTCCCGGAGTCAAGCTCTCCATCCATCCCTCTTTTAA 1502

Db	843	AAGGAAAGAAAGATCTAAAGAACCCACTGTGGGTGAGGGGAAATGGGACACAGAGACATA	902
Qy	633	TGGGCAAGCTCTGCAGGACAGACAGACAAACCCCTGTATCTATGAAGTCTGTGAG	752
Db	903	TGGGCAAGCTCTGCAGGACAGACAGACAAACCCCTGTATCTATGAAGTCTGTGAG	962
Qy	753	GGCAGAGGGAGCAGGAGACCTGGAAACCTCTTGGCCAAAGGGAGTGGGAGACAGAGGGA	812
Db	963	GGCAAGGGGACAGGAGACCTGGAAACCTCTTGGCCAAAGGGAGTGGGAGAGAGGGA	1022
Qy	813	AGGTACAGGCGAAGGGTGTCTATCTTAAGTGGAACTTAATTCGCCAGAGGGCTCAGCAAGCC	872
Db	1023	AGGTACAGGCGAAGGGTGTCTATCTTAAGTGGAACTTAATTCGCCAGAGGGCTCAGCAAGCC	1082
Qy	873	AAGAGGAGACAGCCGTGACGGTAACTTCCCTTACACAGCCCTCAACCCCAACCCAGC	932
Db	1083	AAGAGGAGACAGCCGTGACGGTAACTTCCCTTACACAGCCCTCAACCCCAACCCAGC	1142
Qy	933	GAGCAGGCTGCTGTCCACCCCTGTGCCCAACGACGTGGCTGTGCCAGGGCAAGCCAT	992
Db	1143	GAGCAGGCTGCTGTCCACCCCTGTGCCCAACGACGTGGCTGTGCCAGGGCAAGCCAT	1202
Qy	993	GCCACATCTGTATATAGATGGGGTTTTCCAATAGGTGGTGGTGAATAACGTCAAGA	1052
Db	1203	GCCACATCTGTATATAGATGGGGTTTTCCAATAGGTGGTGGTGAATAACGTCAAGA	1262
Qy	1053	AACTCCTGCTCCGTCCGACCTGTCTGGGGCTTCCAGGCGAAGCGCAAGTGGGGTGGGGGTG	1112
Db	1263	AACTCCTGCTCCGTCCGACCTGTCTGGGGCTTCCAGGCGAAGCGCAAGTGGGGTGGGGGTG	1322
Qy	1113	GGGCTGTGTCTTCTTCCCTCCCAAGGCTGTGTCTNTGGGGCTGTCCCAATGCAACAGG	1172
Db	1323	GGGCTGTGTCTTCTTCCCTCCCAAGGCTGTGTCTNTGGGGCTGTCCCAATGCAACAGG	1382
Qy	1173	ATCACCCTTAACAGAGATGGAAGCGAGGGGATGGATGGGGCTTGGGTCTCCAGAGTTGAGC	1232
Db	1383	ATCACCCTTAACAGAGATGGAAGCGAGGGGATGGATGGGGCTTGGGTCTCCAGAGTTGAGC	1442
Qy	1233	CCGACCTTCTTGCCACTTTCCTCCCTCC-GGCAGTCACTCTCCATCATCCTCCCTCTTTAA	1293
Db	1443	CCGACCTTCTTGCCACTTTCCTCCCTCCGCGGACGTCACTCTCCATCATCCTCCCTCTTTAA	1502
Qy	1292	TCTATGATCTAATAGGAGCTGGGTGTGTGAACA-AACACCCCTAATCGTGTCTTCAAT	1356
Db	1503	TCTATGATCTAATAGGAGCTGGGTGTGTGAACAACACCCCTAATCGTGTCTTCAAT	1562
Qy	1351	ACTCAGACTTAACCATTTGTTGAGGCCAATTCAGAGCTTCTCAATCAGATTTACAATC	1410
Db	1563	ACTCAGACTTAACCATTTGTTGAGGCCAATTCAGAGCTTCTCAATCAGATTTACAATC	1622
Qy	1411	TCCATTTTCATTAACGGGGAAACATCCCCGAGCCACTAGTGTGTGTCTTGTCACTGAA	1470
Db	1623	TCCATTTTCATTAACGGGGAAACATCCCCGAGCCACTAGTGTGTGTCTTGTCACTGAA	1682
Qy	1471	GATTAGATCTGAACCCAGGAGTGTCAACNGCTGTCTCAACTCCCACTCTGGGACATGA	1530
Db	1683	GATTAGATCTGAACCCAGGAGTGTCAACNGCTGTCTCAACTCCCACTCTGGGACATGA	1742
Qy	1531	GGAATTTTCCCTCATTTCTACCTCTCTAAGGCTATGACCCCTCCAGGTCTTCCAGC	1590
Db	1743	GGAATTTTCCCTCATTTCTACCTCTCTAAGGCTATGAC-CCTCCCCAGCCTTCCAGC	1801
Qy	1591	TGGGGGATGGGGGAGCTATAGGAAACCCCACTCTCCATCTGGGATATAGGACCTTCC	1650
Db	1802	TGGGGGATGGGGGAGCTATAGGAAACCCCACTCTCCATCTGGGATATAGGACCTTCC	1861
Qy	1651	ATCAGCCCTTAACCTCTGGGAAATGCTGTGCCCCCAGTACGACTTGTGTGTCTCCAC	1710
Db	1862	ATCAGCCCTTAACCTCTGGGAAATGCTGTGCCCCCAGTACGACTTGTGTGTCTCCAC	1921
Qy	1711	ATACGAAGCAGGGTGGAGGGGAAAGGGTGGGTCTTAGTTAGCAGGGGTCTCCAGGCAAG	1770
Db	1922	ATACGAAGCAGGGTGGAGGGGAAAGGGTGGGTCTTAGTTAGCAGGGGTCTCCAGGCAAG	1981

QY	1771	1882	1831	2042	QY	1890	2102	
QY	CAACCTCCCTCCCTCCATGACCTCTCTGATGATGCTTAAAGGATGGACCTCTCACTGCCA	TCACCTCCCTCCCTCCATGACCTCTCTGATGATGCTTAAAGGATGGACCTCTCACTGCCA	CCAATCTCGGG-CCCTTGGGGGAGACCTGGGAGGGGGCCCTGGAGAGACCTGACGCTGG	CCAATCTCGGGCCCTTGGGGAGAGACCTGGGAGGGGGCCCTGGAGAGACCTGACGCTGG	CAACTGTATACACATTAAGGACAGCTTCCACAGA	CAACTGTATACACATTAAGGACAGCTTCCACAGA	CAACTGTATACACATTAAGGACAGCTTCCACAGA	CAACTGTATACACATTAAGGACAGCTTCCACAGA
Db	1882	1831	2042	1890	2102	1882	1831	

Search completed: October 24, 2004, 13:54:30
Job time : 8565 secs

QY 1709 ACATACAGAGAGGCTGAGGAGGAGGAGGCTGCTCACTTACAGAGGGTCCCGAGGCA 1768
DB 61 AATATAGAGAGAGGCTGAGGAGGAGGCTGCTCACTTACAGAGGGTCCCGAGGCA 120
QY 1769 AGTACAGGCTCTCCCTCCCACTGCTCTCTGCTGAGTGTGCTTAAAGGTGGCCCTCACTCC 1828
DB 121 AGTACAGGCTCTCTCCCTCCCACTGCTCTCTGCTGAGTGTGCTTAAAGGTGGCCCTCACTCC 180
QY 1829 CACCACTCTGGG-CCCTTGGGGAGAGACTGGGAGAGGGGCGCTGGAGAGAGCCCTGAGCT 1887
DB 181 CACCACTCTGGGCGCCCTTGGGGAGAGAGACTGGGAGAGGGGCGCTGGAGAGAGCCCTGAGCT 240
QY 1888 GGAACC 1893
DB 241 GGAACC 246

RESULT 2

US-09-620-312D-669
Sequence 669, Application US/09620312D
Patent No. 656962
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyang
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yunding
APPLICANT: Wang, Dunhui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast
APPLICANT: Dmanac, Radoje T.
TITLE OF INVENTION: No. 656962el Nucleic Acids and
FILE REFERENCE: Polypeptides
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pc_fl_genes Version 1.0
SEQ ID NO 669
LENGTH: 719
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (81)..(578)
US-09-620-312D-669

Query Match 8.6%; Score 171.4; DB 4; Length 719;
Best Local Similarity 78.2%; Pred. No. 9.7e-41;
Matches 287; Conservative 0; Mismatches 61; Indels 19; Gaps 6;

QY 33 GCCCATCCCAACCAAGCCCGGTGGGCAAGCAATCAAGACCTTAGAGAGAGCCCTATGA 92
DB 254 GGCCTTCCCAAGCCCGCGGTGGGCAAGCAATCAAGACCTTAGAGAGAGCCCTATGA 313
QY 93 GTTTGGGTGAGAGTGAAGACTTCTCACTTGAAGACATCTTGTGACCACTCCAAACA 152
DB 314 GTTTGGGTGAGAGTGAAGACTTCTCACTTGAAGACATCTTGTGACCACTCCAAACA 370
QY 153 AACCAACATCAGTGTGGGGCTGAGAAAGCTGGGGCTGAACGACACTGTACATGAAA 212

DB 371 CAACCAATCAGAGTGGG--CTGAGAGCTGGCGGCTGACGGCACTGTATGAACACC 428
QY 213 CACCTTACGACTACACAGTGCAGACTGGCGGAGAGCGTGAACCCGAGCGTGTGA 272
DB 429 TTGGCT-----CACAAGTCCAG--CTGCCGAGAGACGTGAGACC--GACGTGCTGA 477
QY 273 CTTGGTACTGTGCGGAGAGACGACAGCTCACTGTATCCGGGACAGCGCTACCCGAGTA 332
DB 478 CTTGG--CTTGGCGGAGAGAGAGAGCTTCACTATCCGGGACAGCGCTACCCGAGTA 534
QY 333 CAGACACGTCCAGACAGACCTTCCGAGAGAGATCAAAATCTGAGTCCCTTCC 392
DB 535 CAGACACGTCCAGACAGACCTTCCGAGAGAGATCAAAATCTGAGTCCCTTCC 594
QY 393 TTTCCT 399
DB 595 TTTCCT 601

RESULT 3

US-09-513-999C-20093
Sequence 20093, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclet, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
Patent No. 6783961
FILE REFERENCE: 59.US2.REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 20093
LENGTH: 125
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 120
OTHER INFORMATION: m=a or c
US-09-513-999C-20093

Query Match 6.0%; Score 119.6; DB 4; Length 125;
Best Local Similarity 98.3%; Pred. No. 9e-26;
Matches 119; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1398 CAGATTACATCTCCATTTTCATTAAAGGGGAAACATCCCGAGCGCACTGAGTGTG 1457
DB 1 CAGATTACATCTCCATTTTCATTAAAGGGGAAACATCCCGAGCGCACTGAGTGTG 60
QY 1458 CTTTGCATGAGAGTGAATCGAACCGAGGTGCAACNGTGTCTCACTCCGAC 1517
DB 61 CTTTGCATGAGAGTGAATCGAACCGAGGTGCAACNGTGTCTCACTCCGAC 120

QY 1518 C 1518
DB 121 C 121

RESULT 4

US-09-513-999C-2322
Sequence 2322, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclet, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

Patent No. 6281344
GENERAL INFORMATION:
APPLICANT: Szoetax, Jack W.
APPLICANT: Roberts, Richard W.
APPLICANT: Liu, Rihue
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
TITLE OF INVENTION: FUSIONS
FILE REFERENCE: 00786/350007
CURRENT APPLICATION NUMBER: US/09/244,796
CURRENT FILING DATE: 1999-02-05
EARLIER APPLICATION NUMBER: 60/035,963
EARLIER FILING DATE: 1997-01-27
EARLIER APPLICATION NUMBER: 60/064,491
EARLIER FILING DATE: 1997-11-06
EARLIER APPLICATION NUMBER: 09/007,005
EARLIER FILING DATE: 1998-01-14
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE: Translation template
OTHER INFORMATION: Translation template
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(289)
OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match 2.1%; Score 42.6; DB 3; Length 289;
Best Local Similarity 3.8%; Pred. No. 0.011;
Matches 8; Conservative 101; Mismatches 100; Indels 0; Gaps 0;
QY 627 TGGGTCAGGGAAGATCTAAAGAACCCAGCTGTGGTCAGGGAATGGACCCAG 696
DB 42 URTDRCRARRUTGRNRRNRNRRNRNRRNRNRRNRNRRNRNRRNRNRRNR 101
QY 687 GACATATGGCAGCCTGTGACAGACAGACAGACAGACAAACCTGTATGAGCT 746
DB 102 NRSNRNRRNRNRRNRNRRNRNRRNRNRRNRNRRNRNRRNRNRRNRNRRNR 161
QY 747 CTGAGGCGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 806
DB 162 NRSNRNRRNRNRRNRNRRNRNRRNRNRRNRNRRNRNRRNRNRRNRNRRNR 221
QY 807 GAGGAGGTCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 835
DB 222 ARGRCRURGRGRURRARRCRURURU 250

RESULT 10
US-09-962A-6/c
Sequence 6, Application US/0990962A
Patent No. 6534272
GENERAL INFORMATION:
APPLICANT: POLYCHRONAKOS, Constantin
APPLICANT: VAFIADIS, Petros
APPLICANT: GRABIS, Rosemarie
APPLICANT: OUNISSI-BENKALHA, Houria
TITLE OF INVENTION: DNA ASSAY FOR THE PREDICTION OF
TITLE OF INVENTION: AUTOIMMUNE DIABETES
FILE REFERENCE: 1770-213US-1 FC
CURRENT APPLICATION NUMBER: US/09/909,962A
CURRENT FILING DATE: 2001-07-23
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 2130
TYPE: DNA
ORGANISM: human
US-09-909-962A-6

Query Match 2.0%; Score 39.8; DB 4; Length 2130;
Best Local Similarity 48.2%; Pred. No. 0.22;
Matches 145; Conservative 0; Mismatches 152; Indels 4; Gaps 1;
QY 257 CCCGAGCGTGGGTGACCTCGGCTACTGTGGAGAGAGGAGAGGAGGAGGAGGAG 316
DB 2130 CCCGAGCGCTGTCCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2071
QY 317 AGCGGTACCCCGATACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376
DB 2070 CCCGAGCGCTGTCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2011
QY 377 GTCCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTT 436
DB 2010 ACACCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 1951
QY 437 ACCCATTAACAAGC---TCAGAGACATCTCAGCCAGGTTCTAGCCCCCAGCACC 492
DB 1950 CCCCTGTCCCGACACCCCTGTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAG 1891
QY 493 CAGACCCGAGTGAACATCTCCCAACTAGAGGAGGAGGAGGAGGAGGAGGAGGAG 552
DB 1890 CTGTCCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1831
QY 553 G 553
DB 1830 G 1830

RESULT 11
US-09-962A-7/c
Sequence 7, Application US/0990962A
Patent No. 6534272
GENERAL INFORMATION:
APPLICANT: POLYCHRONAKOS, Constantin
APPLICANT: VAFIADIS, Petros
APPLICANT: GRABIS, Rosemarie
APPLICANT: OUNISSI-BENKALHA, Houria
TITLE OF INVENTION: DNA ASSAY FOR THE PREDICTION OF
TITLE OF INVENTION: AUTOIMMUNE DIABETES
FILE REFERENCE: 1770-213US-1 FC
CURRENT APPLICATION NUMBER: US/09/909,962A
CURRENT FILING DATE: 2001-07-23
NUMBER OF SEQ ID NOS: 7
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 2130
TYPE: DNA
ORGANISM: human
US-09-909-962A-7

Query Match 2.0%; Score 39.8; DB 4; Length 2130;
Best Local Similarity 48.2%; Pred. No. 0.22;
Matches 145; Conservative 0; Mismatches 152; Indels 4; Gaps 1;

QY 257 CCCGAGCGTGGGTGACCTCGGCTACTGTGGAGAGAGGAGAGGAGGAGGAGGAG 316
DB 2130 CCCGAGCGCTGTCCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2071
QY 317 AGCGGTACCCCGATACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376
DB 2070 CCCGAGCGCTGTCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2011
QY 377 GTCCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTT 436
DB 2010 ACACCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 1951
QY 437 ACCCATTAACAAGC---TCAGAGACATCTCAGCCAGGTTCTAGCCCCCAGCACC 492
DB 1950 CCCCTGTCCCGACACCCCTGTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1891
QY 493 CAGACCCGAGTGAACATCTCCCAACTAGAGGAGGAGGAGGAGGAGGAGGAGGAG 552

Mon Oct 25 11:39:41 2004

us-09-977-418-7.rn1

Page 7

	Matches	90;	Conservative	0;	Mismatches	86;	Indels	0;	Gaps	0;
QY	366	TCAAAATCTGAGTACCTCTCTCCCTTCCCTTCCCTGTCGCCGCGCCACGCTGCGACGA	425							
Db	311	TCAAAATCTGAGTACCTCTCTCCCTTCCCTTCCCTGTCGCCGCGCCACGCTGCGACGA	252							
QY	426	AAGCCTGCTAACCCCATTAACAAGCTCCAGGACATCTCAGCCCAAGTTCTAGCCCCA	485							
Db	251	CC	192							
QY	486	CGACCCCCAGACCCGAGTGAGCCATCTCTCCAACTAGGGCCCTCCACTCTATCC	541							
Db	191	CC	136							

Search completed: October 24, 2004, 15:44:03
Job time : 179 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 23, 2004, 21:13:22 ; Search time 979 Seconds
(without alignments)
10654.341 Million cell updates/sec

Title: US-09-977-418-7
Perfect score: 1987
Sequence: 1 tgaatgaagcctgtccatg.....cnnnctacttgagagaaa 1987

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 413486 segs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Geneseg_235004:*
2: geneseg_1980s:*
3: geneseg_1990s:*
4: geneseg_2000a:*
5: geneseg_2001b:*
6: geneseg_2002a:*
7: geneseg_2002b:*
8: geneseg_2003a:*
9: geneseg_2003b:*
10: geneseg_2003c:*
11: geneseg_2003d:*
12: geneseg_2004a:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1979	99.6	1987	AAf27852	AAf27852 Human NOV
2	1979	99.6	1987	ACD40257	ACD40257 Human bre
3	1979	99.6	1987	ADf66734	ADf66734 CDNA enco
4	1979	99.6	1987	ADf66734	ADf66734 Human NOV
5	1975	99.4	1987	ADf19771	ADf19771 Human NOV
6	1974	99.4	1987	ADf56372	ADf56372 Human CN
7	1684	84.8	2192	AAZ33583	AAZ33583 Human bre
8	1670	84.0	2142	AAI60565	AAI60565 Human pol
9	1670	84.0	2142	ADc31919	ADc31919 Human nov
10	1668	83.9	2220	ADm19231	ADm19231 Novel hum
11	1666	83.6	3503	ADg22267	ADg22267 Human sof
12	1660	83.6	2410	ADb62818	ADb62818 Human CN
13	1658	83.5	2701	AAK94832	AAK94832 Human ful
14	1658	83.5	2701	ADL31950	ADL31950 Full leng
15	1648	82.9	2783	ADc29984	ADc29984 Human nov
16	1647	82.9	2152	ADQ17578	ADQ17578 Human sof
17	1627	81.7	3490	ACC46692	ACC46692 Human dit
18	1612	81.2	4056	ADM20171	ADM20171 Alternati
19	1084	54.6	1550	AAf54984	AAf54984 Nucleotid
20	439	22.1	514	AAK94049	AAK94049 Human CN
21	439	22.1	514	ADL30476	ADL30476 3' end of

22	414.2	20.8	421	9	ACH47257	ACH47257 Human inf
23	410.4	20.7	412	9	ACH18011	ACH18011 Human adu
24	409.4	20.6	412	9	ACH17535	ACH17535 Human adu
25	402	20.2	402	6	ABL62195	ABL62195 Colon ade
26	402	20.2	402	6	ABK64128	ABK64128 Human ben
27	399.4	20.1	486	9	ACH16596	ACH16596 Human adu
28	384.2	19.3	757	12	AAK93823	AAK93823 Human CN
29	384.2	19.3	757	12	ADL30250	ADL30250 3' end of
30	375	18.9	400	9	ACH16935	ACH16935 Human adu
31	375	18.9	400	9	AAI12114	AAI12114 Probe #20
32	375	18.9	458	4	AAK53819	AAK53819 Human foe
33	375	18.9	458	4	AAI33455	AAI33455 Probe #21
34	375	18.9	458	4	ABK43376	ABK43376 Human bre
35	375	18.9	458	4	ABK23567	ABK23567 Probe #20
36	375	18.9	458	4	AAK27535	AAK27535 Human bon
37	375	18.9	458	4	AAK02080	AAK02080 Human bra
38	375	18.9	458	4	ABK27102	ABK27102 Human liv
39	375	18.9	458	5	AAI02035	AAI02035 Probe #20
40	375	18.9	458	6	ABK92850	ABK92850 Human gen
41	326.2	16.4	368	4	AAK92850	AAK92850 Human CN
42	326.2	16.4	368	12	ADL29377	ADL29377 3' end of
43	277.8	14.0	305	4	AAf27901	AAf27901 Human NOV
44	277.8	14.0	305	9	ACD40305	ACD40305 Breast tu
45	277.8	14.0	305	11	ADM56439	ADM56439 Human CN

ALIGNMENTS

RESULT 1	AAf27852	standard; CDNA; 1987 BP.
ID	AAf27852	
AC	AAf27852	
XX	30-MAR-2001	(first entry)
DT	XX	Human NOV4 CDNA.
DE	XX	
XX	Human; NOVX; antiinflammatory; cyrostatic; neuroprotective;	
KW	cerebroprotective; immunomodulator; vulnerary; vasotrophic; gene therapy;	
KW	hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;	
KW	diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease; ss.	
OS	Homo sapiens.	
XX	WO200075321-A2.	
PN	14-DEC-2000.	
PD	XX	
XX	01-JUN-2000; 2000WO-US015303.	
PF	XX	
XX	03-UTN-1999; 99US-0137322P.	
PR	16-MAR-2000; 2000US-0189810P.	
PR	22-MAR-2000; 2000US-0191158P.	
PR	30-MAR-2000; 2000US-0193086P.	
PR	31-MAR-2000; 2000US-00137322.	
XX	(CURA-) CURAGEN CORP.	
PA	XX	
XX	Shinketsu RA, Fernandes E, Herrman J, Vernet C;	
PI	XX	
DR	WPI; 2001-102403/11.	
XX	P-PSDB; AAB61132.	
PT	New NOVX polypeptides and polynucleotides, useful in gene therapy, as a	
PT	diagnostic marker, protein therapeutic, antibody or small molecule drug	
PT	target for treating immune, proliferative and metabolic diseases and	
XX	wound healing.	
XX	Claim 8; Page 26-27; 194pp; English.	
PS	The present sequence encodes a new isolated polypeptide (NOVX). The NOVX	
XX		
CC		

CC polypeptides, NOXV nucleic acids, and anti-NOXV antibodies are useful for
 CC treating or preventing NOXV-associated disorders. They are also useful
 CC for determining the presence of or a predisposition to a disease
 CC associated with altered levels of the NOXV polypeptide or nucleic acid.
 CC These NOXV-associated disorders include hyperplasia, tumours,
 CC rheumatoid arthritis, Dupuytren's contracture, diabetic complications,
 CC oedema, senile dementia or Alzheimer's disease. The NOXV polypeptides
 CC are especially useful in gene therapy. Specifically, NOXV is useful as a
 CC diagnostic marker or prognostic marker, protein therapeutic and antibody
 CC target or small molecule drug target to treat disorders in the immune
 CC response pathway, thyroid and metabolic diseases, bone metabolic
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive
 CC disorders), proliferative diseases, or tissue regeneration and
 CC development (e.g. wound healing or treatment of burns)

XX Sequence 1987 BP; 448 A; 613 C; 532 G; 386 T; 0 U; 8 Other;

Query Match 99.6%; Score 1979; DB 4; Length 1987;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 TGAATGAGGCTGTCCATGCTCTGTAAGAGCCCATCCCAAGCCCGGTGGCA 60
 1 TGAATGAGGCTGTCCATGCTCTGTAAGAGCCCATCCCAAGCCCGGTGGCA 60
 QY 61 GGGACATCAAGACCTTAGAGAGCGCTATGAGTTGGGAGAGAGACTTCA 120
 DB 61 GGGACATCAAGACCTTAGAGAGCGCTATGAGTTGGGAGAGAGACTTCA 120
 QY 121 CCTGAAGACATCTTGTCAACCTCCAAACCAACCAACATGAGGTGGGGCTGAGA 180
 DB 121 CCTGAAGACATCTTGTCAACCTCCAAACCAACCAACATGAGGTGGGGCTGAGA 180
 QY 181 AGTGGCGGCTGAACGGAACCTGTATATAAACCCTTACACTACAGAGTGGCACT 240
 DB 181 AGTGGCGGCTGAACGGAACCTGTATATAAACCCTTACACTACAGAGTGGCACT 240
 QY 241 GCGGAGAGAGTGAACCCGAGCGTGGCTGACCTGCTGCTGCTGAGGAGAGAG 300
 DB 241 GCGGAGAGAGTGAACCCGAGCGTGGCTGACCTGCTGCTGCTGAGGAGAGAG 300
 QY 301 CCTCACTAGTCCGGGACCGCTGACCCGCTACAGAACAGCTCCAGAGACTTCCGAG 360
 DB 301 CCTCACTAGTCCGGGACCGCTGACCCGCTACAGAACAGCTCCAGAGACTTCCGAG 360
 QY 361 GGAATCAAAATCTGAGTGCCTCTGCTTCCCTTCCCTGCTCCCGCCCAAGGCTGC 420
 DB 361 GGAATCAAAATCTGAGTGCCTCTGCTTCCCTTCCCTGCTCCCGCCCAAGGCTGC 420
 QY 421 GAGCAAAAGCTGCTAACCCCATTAACAAGCTCCAGAGCACTCAGCCAGGTTCTAGC 480
 DB 421 GAGCAAAAGCTGCTAACCCCATTAACAAGCTCCAGAGCACTCAGCCAGGTTCTAGC 480
 QY 481 CCCCAAGCAACCCCAAGTGAACATCTCTCCCAAACTAAGGAGCTTCACTATC 540
 DB 481 CCCCAAGCAACCCCAAGTGAACATCTCTCCCAAACTAAGGAGCTTCACTATC 540
 QY 541 CAGGGAGGCTCAGGAGCTCCCTGAGCTGACATGATGCCAGATTTCAGATTGGGCTC 600
 DB 541 CAGGGAGGCTCAGGAGCTCCCTGAGCTGACATGATGCCAGATTTCAGATTGGGCTC 600
 QY 601 CCTCACTTATCAAGATCAAGGAGCTGGGGTCAAGGAGAGATCTTAAAGAACCACT 660
 DB 601 CCTCACTTATCAAGATCAAGGAGCTGGGGTCAAGGAGAGATCTTAAAGAACCACT 660
 QY 661 GTGGGTCAAGGAGATGGGACAGAGCATTTGGGAAAGCTTCCAGAGACAGAGCA 720
 DB 661 GTGGGTCAAGGAGATGGGACAGAGCATTTGGGAAAGCTTCCAGAGACAGAGCA 720
 QY 721 GACAAACCTCTGATCTATGAAGTCTTGCAGAGGAGAGGAGACCTGAGACCTT 780
 DB 721 GACAAACCTCTGATCTATGAAGTCTTGCAGAGGAGAGGAGACCTGAGACCTT 780

QY 781 CTTGGCAAGGAGAGTGGAGAGACAGAGGAGAGTCAAGGAGAGGCTCTATCTAAG 840
 DB 781 CTTGGCAAGGAGAGTGGAGAGACAGAGGAGAGTCAAGGAGAGGCTCTATCTAAG 840
 QY 841 TGAATCAATTTGCCCGAGGGCTCAGCAAGCCCAAGAGAGACAGGCTGACGTTAACTT 900
 DB 841 TGAATCAATTTGCCCGAGGGCTCAGCAAGCCCAAGAGAGACAGGCTGACGTTAACTT 900
 QY 901 CCCCTTACAGAGCTTCAAGCCCAAGCCAGAGAGAGGCTGCTGCCACCCCGAGGCC 960
 DB 901 CCCCTTACAGAGCTTCAAGCCCAAGCCAGAGAGAGGCTGCTGCCACCCCGAGGCC 960
 QY 961 CCAGCAGCTGCTGTGCGCAGGAGAGCCATCCATCTGTATATAGTGGGTTT 1020
 DB 961 CCAGCAGCTGCTGTGCGCAGGAGAGCCATCCATCTGTATATAGTGGGTTT 1020
 QY 1021 CCAATACAGCTGCTGTGAGAAACCTGCATGAAACCTCCGCTCCGCGCTGAGGG 1080
 DB 1021 CCAATACAGCTGCTGTGAGAAACCTGCATGAAACCTCCGCTCCGCGCTGAGGG 1080
 QY 1081 CCTTCAGGCAAGGCAAGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGG 1140
 DB 1081 CCTTCAGGCAAGGCAAGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGG 1140
 QY 1141 TGTGTTCTGAGGCTGCTCCATGAGAGAGAGTCACTAAGAGATGGAAGCCAGGGC 1200
 DB 1141 TGTGTTCTGAGGCTGCTCCATGAGAGAGAGTCACTAAGAGATGGAAGCCAGGGC 1200
 QY 1201 ATGATGAGGCTTTGGGCTCTGAGGTTGAGACCCAGCTTCTTCCACCTTCCCTCCG 1260
 DB 1201 ATGATGAGGCTTTGGGCTCTGAGGTTGAGACCCAGCTTCTTCCACCTTCCCTCCG 1260
 QY 1261 CAGTCACTTCCATCTCATCCCTCTTATATATGAACTCTATGAGGCTGCTGTGTA 1320
 DB 1261 CAGTCACTTCCATCTCATCCCTCTTATATATGAACTCTATGAGGCTGCTGTGTA 1320
 QY 1321 ACAACACACCCCTATGCTTCTCTTCAATACTCAGCATTTACATGTTGAGGCCAAT 1380
 DB 1321 ACAACACACCCCTATGCTTCTCTTCAATACTCAGCATTTACATGTTGAGGCCAAT 1380
 QY 1381 TCAAGAGCTTTCATAATCAAGATTACATCTCATTTTCAATTAACGGGAAACATCCCG 1440
 DB 1381 TCAAGAGCTTTCATAATCAAGATTACATCTCATTTTCAATTAACGGGAAACATCCCG 1440
 QY 1441 AGCAGCTGAGTGTGCTTGTGCACTGAAGGTTAATGAAACCCAGGAGTCAACGCG 1500
 DB 1441 AGCAGCTGAGTGTGCTTGTGCACTGAAGGTTAATGAAACCCAGGAGTCAACGCG 1500
 QY 1501 TGTCTCAACTCCCACTCTGGGCACTGAGGAGATTTCCCTCATTTCACTCTTAA 1560
 DB 1501 TGTCTCAACTCCCACTCTGGGCACTGAGGAGATTTCCCTCATTTCACTCTTAA 1560
 QY 1561 GGGTATGACCCCTCCCAAGTCTTCCAGTGGGAGTGGGAGGATCAAGGAAAGCC 1620
 DB 1561 GGGTATGACCCCTCCCAAGTCTTCCAGTGGGAGTGGGAGGATCAAGGAAAGCC 1620
 QY 1621 CCCATCTCCATCTGGAGATGAGGACCTTCCATCAGGCTTAAACCTGGGAAATGCTGCTG 1680
 DB 1621 CCCATCTCCATCTGGAGATGAGGACCTTCCATCAGGCTTAAACCTGGGAAATGCTGCTG 1680
 QY 1681 CCCCAAGTACTTGTGTTTCTGCTTCCCATTAACAAGAGAGGAGAGGAGGAGG 1740
 DB 1681 CCCCAAGTACTTGTGTTTCTGCTTCCCATTAACAAGAGAGGAGAGGAGGAGG 1740
 QY 1741 GTCTCAGTTAGAGGAGTCCCAAGGAGAGTCAAGCTCTCTCCATGAGCTCTGAGTC 1800
 DB 1741 GTCTCAGTTAGAGGAGTCCCAAGGAGAGTCAAGCTCTCTCCATGAGCTCTGAGTC 1800
 QY 1801 AGTGTCTTTAGAGGAGTCTCTCACTCCCACTCTGAGGAGGAGAGAGAGGAG 1860
 DB 1801 AGTGTCTTTAGAGGAGTCTCTCACTCCCACTCTGAGGAGGAGAGAGAGGAG 1860

QY 1661 AGGGGGCCGCTGGAGAGAGCCCTGACCGCTTGAACCTGTATACACAAATTAAGGACAGCTCAG
 Db 1661 AGGGGGCCGCTGGAGAGAGCCCTGACCGCTTGAACCTGTATACACAAATTAAGGACAGCTCAG 1920
 QY 1921 AGACNTCTGGAGAGCGCCCTGCGCNGAGATTCTTCAAACTTAGGCGAGGGCGNNNNCTTACTTG 1980
 Db 1921 AGACNTCTGGAGAGCGCCCTGCGCNGAGATTCTTCAAACTTAGGCGAGGGCGNNNNCTTACTTG 1980
 QY 1981 AGAGAAA 1987
 Db 1981 AGAGAAA 1987
 RESULT 2
 ACD40257
 ID ACD40257 standard; DNA; 1987 BP.
 XX ACD40257;
 AC ACD40257;
 XX
 DT 03-SEP-2003 (first entry)
 DE Human breast tumour associated protein 47-like polypeptide NOV4 DNA.
 XX
 XX Tissue typing; cancer; breast cancer; colon cancer; lung cancer; sarcoma;
 KW pancreatic cancer; uterine cancer; organ transplantation disorder; ds;
 KW cardiovascular disease; melanoma; atherosclerosis; diabetes mellitus;
 KW ischaemic heart disease; haemorrhage; peripheral vascular disease; gene;
 KW thrombosis; hypertension; systemic lupus erythematosus; haematopoiesis;
 KW tissue regeneration; wound healing; hyperproliferative disorder;
 KW psoriasis; neural disorder; Parkinson's disease; Alzheimer's disease;
 KW Huntington's disease; multiple sclerosis; amyotrophic lateral sclerosis;
 KW ALS; peripheral neuropathy; nervous system tumour; neurotoxin; tremor;
 KW neuropathy; acute brain injury; peripheral nerve trauma; human; NOV4;
 KW gene therapy; epilepsy; breast tumour associated protein 47.
 XX
 XX Homo sapiens.
 OS
 XX US2003027158-A1.
 PN
 PD 06-FEB-2003.
 XX
 PF 15-OCT-2001; 2001US-00977418.
 XX
 XX 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX
 XX (CDBA-) CDBAGEN CORP.
 PA
 PI Shimkets RA, Fernandes E, Herrman J, Vernet C;
 XX
 XX WPI: 2003-492028/46.
 DR P-PSDB; ABO23236.
 PT
 PT New nucleic acid sequence encoding a human breast tumor-associated
 PT protein 47-like polypeptide, useful for treating cardiovascular
 PT disorders, neural disorders, diabetes mellitus and cancers.
 PS
 PS Claim 8; Page 16-17; 100pp; English.
 CC The invention relates to a new isolated NOV4 nucleic acid. The nucleic
 CC acid is useful for identifying a compound that binds the nucleic acid.
 CC The nucleic acid is useful in gene therapy, in screening assays, in
 CC detection assays e.g. chromosomal mapping, cell and tissue typing and
 CC forensic biology, predictive medicine e.g. diagnostic assays, prognostic
 CC assays, monitoring clinical trials and pharmacogenomics and methods of
 CC treatment including therapeutic and prophylactic. The nucleic acid is
 CC also useful for expressing NOV4 protein. The nucleic acid is also useful
 CC to provide polynucleotide reagents e.g. labelled probes that are useful
 CC in an in situ hybridisation technique, for identifying a specific tissue

Query Match	Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches	0;	Indels	0;	Gaps	0
Matches 1987;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0	
Sequence 1987 BP; 448 A; 513 C; 532 G; 386 T; 0 U; 8 Other;									
1	TGATTAAGGCGCTGTCATAGCTCGTCAAAAGCCCATCCCAACCAAGCCCGGTGAGCA	60							
1	TGATTAAGGCGCTGTCATAGCTCGTCAAAAGCCCATCCCAACCAAGCCCGGTGAGCA	60							
1	TGATTAAGGCGCTGTCATAGCTCGTCAAAAGCCCATCCCAACCAAGCCCGGTGAGCA	60							
61	GCGACATCAAGACCCCTTAGAGAGACGCTTAGAGTTTGGCGTGCAGTGAAGACTTCTCA	120							
61	GCGACATCAAGACCCCTTAGAGAGACGCTTAGAGTTTGGCGTGCAGTGAAGACTTCTCA	120							
61	GCGACATCAAGACCCCTTAGAGAGACGCTTAGAGTTTGGCGTGCAGTGAAGACTTCTCA	120							
121	CCGGAAGACATCATTTGTACCAACCTCCGCAACAAACCAACATCGAGGTGCGGGGTGAGAA	180							
121	CCGGAAGACATCATTTGTACCAACCTCCGCAACAAACCAACATCGAGGTGCGGGGTGAGAA	180							
121	CCGGAAGACATCATTTGTACCAACCTCCGCAACAAACCAACATCGAGGTGCGGGGTGAGAA	180							
181	AGCTGCGGCTGAAACGACACTGTACATGAAGAACCTTAGACATACCAAGTGCAGACT	240							
181	AGCTGCGGCTGAAACGACACTGTACATGAAGAACCTTAGACATACCAAGTGCAGACT	240							
181	AGCTGCGGCTGAAACGACACTGTACATGAAGAACCTTAGACATACCAAGTGCAGACT	240							
241	GCGGAGAGACGTGACACCCGAGCGGTGACCTCGGCTACTCTGGGAGAGACGAGCAG	300							
241	GCGGAGAGACGTGACACCCGAGCGGTGACCTCGGCTACTCTGGGAGAGACGAGCAG	300							
241	GCGGAGAGACGTGACACCCGAGCGGTGACCTCGGCTACTCTGGGAGAGACGAGCAG	300							
301	CTTCACTAGTCCGGGACGCGTCAACCCGATACAGAAACAGTCCAGACAGACTTCCGGAC	360							
301	CTTCACTAGTCCGGGACGCGTCAACCCGATACAGAAACAGTCCAGACAGACTTCCGGAC	360							
301	CTTCACTAGTCCGGGACGCGTCAACCCGATACAGAAACAGTCCAGACAGACTTCCGGAC	360							
361	GAGATCAAAATCTGAGTGTCTCTCCCTTCCCTTCCCTGTCCCGCCGACGCTGC	420							
361	GAGATCAAAATCTGAGTGTCTCTCCCTTCCCTTCCCTGTCCCGCCGACGCTGC	420							
361	GAGATCAAAATCTGAGTGTCTCTCCCTTCCCTTCCCTGTCCCGCCGACGCTGC	420							
421	CAGCAAGGCTGCTAACCCATTACAAACAGTCCAGACACTTCAAGCCAGGTTTAGC	480							
421	CAGCAAGGCTGCTAACCCATTACAAACAGTCCAGACACTTCAAGCCAGGTTTAGC	480							
421	CAGCAAGGCTGCTAACCCATTACAAACAGTCCAGACACTTCAAGCCAGGTTTAGC	480							
481	CCCCAGCAGCCCGACAGCCAGGTGAGACATCTTCCAACTAGAGGCTTCACTTATC	540							
481	CCCCAGCAGCCCGACAGCCAGGTGAGACATCTTCCAACTAGAGGCTTCACTTATC	540							
481	CCCCAGCAGCCCGACAGCCAGGTGAGACATCTTCCAACTAGAGGCTTCACTTATC	540							
541	CAGGCGAGGCTCAGGACTCTCCCTGCGCTGACACAGATGCCAGATTTCAGATTGGGCTC	600							
541	CAGGCGAGGCTCAGGACTCTCCCTGCGCTGACACAGATGCCAGATTTCAGATTGGGCTC	600							
541	CAGGCGAGGCTCAGGACTCTCCCTGCGCTGACACAGATGCCAGATTTCAGATTGGGCTC	600							
601	CGTCACTTATCCAGAGTACAGGGGCTGGGGTCAAGGAGGAAGATCTTAAAGAACCACT	660							
601	CGTCACTTATCCAGAGTACAGGGGCTGGGGTCAAGGAGGAAGATCTTAAAGAACCACT	660							
601	CGTCACTTATCCAGAGTACAGGGGCTGGGGTCAAGGAGGAAGATCTTAAAGAACCACT	660							
661	GTGGGTCAAGGAGATGGAGACGACAGACATTTGGGCAAGCTCTGACAGACAGACAGACA	720							

Db 661 GTGGGTGAGGGGAAATGGGACCAAGCAGACATAATGGGCAAGCTCTGACAGACAGACAGACA 720
 Qy 721 GACAAACCCCTGATCTATGAAGTCTCTGAGGGGCAAGGGGACCAAGGACCTGGAACCT 780
 Db 721 GACAAACCCCTGATCTATGAAGTCTCTGAGGGGCAAGGGGACCAAGGACCTGGAACCT 780
 Qy 781 CTGGCCAAAGGGAGTGGGAGAGACAGAGGAAAGTCAAGGCAAGGATGCTTATCTAAG 840
 Db 781 CTGGCCAAAGGGAGTGGGAGAGACAGAGGAAAGTCAAGGCAAGGATGCTTATCTAAG 840
 Qy 841 TGAACCTAATTTGGCCAGGGCTCAGAGAGCCCAAGAGGAGACAGACGCTGACCGTAACTT 900
 Db 841 TGAACCTAATTTGGCCAGGGCTCAGAGAGCCCAAGAGGAGACAGACGCTGACCGTAACTT 900
 Qy 901 CCCCTTACAGGCTTCAAGCCCAAGCCAGGAGAGGCTGCTGCCACCCCGTGGCC 960
 Db 901 CCCCTTACAGGCTTCAAGCCCAAGCCAGGAGAGGCTGCTGCCACCCCGTGGCC 960
 Qy 961 CCAGCCAGCTGGTGTGGCCAGGGCAGAGCCATGCTCAATCTGTATATAGATGGGTTTTT 1020
 Db 961 CCAGCCAGCTGGTGTGGCCAGGGCAGAGCCATGCTCAATCTGTATATAGATGGGTTTTT 1020
 Qy 1021 CCATACAGCTGGTGTGAAAAAACTGCATGAAACTCTGCGGCTCTGCGCTGCTGGGG 1080
 Db 1021 CCATACAGCTGGTGTGAAAAAACTGCATGAAACTCTGCGGCTCTGCGCTGCTGGGG 1080
 Qy 1081 CCTCCAGGCAAGCCAGTGGGGTGGGGGTGGGGTGGTCTTCTCTCCCTCCACAGGCC 1140
 Db 1081 CCTCCAGGCAAGCCAGTGGGGTGGGGGTGGGGTGGTCTTCTCTCCCTCCACAGGCC 1140
 Qy 1141 TGTGTCNTGGGGCTGCTCCATGACAGAGATCACTCAACAGAGATGGAAGCCAGGGC 1200
 Db 1141 TGTGTCNTGGGGCTGCTCCATGACAGAGATCACTCAACAGAGATGGAAGCCAGGGC 1200
 Qy 1201 ATGATGAGGAGCTTTGGGTCCTGAGAGTTGAGCCCAAGCTTCTTGGCACCTTCCCTCCGG 1260
 Db 1201 ATGATGAGGAGCTTTGGGTCCTGAGAGTTGAGCCCAAGCTTCTTGGCACCTTCCCTCCGG 1260
 Qy 1261 CAGTACGCTCTCCATCCATCCCTCTTTAATCTAATGAACTATAGCTCGTGTGCTTA 1320
 Db 1261 CAGTACGCTCTCCATCCATCCCTCTTTAATCTAATGAACTATAGCTCGTGTGCTTA 1320
 Qy 1321 ACAACACACCCCTATGTTGCTTCAAAATACAGCATTTACATTTGTTGAGGCCAAT 1380
 Db 1321 ACAACACACCCCTATGTTGCTTCAAAATACAGCATTTACATTTGTTGAGGCCAAT 1380
 Qy 1381 TCAGAGCTTTCTCAATCAGATTATACATCTCCATTTTCAATACGGGAAACATCCCG 1440
 Db 1381 TCAGAGCTTTCTCAATCAGATTATACATCTCCATTTTCAATACGGGAAACATCCCG 1440
 Qy 1441 AGCACTGAGTCTGTGCTTGTCACTGAAAGTTAGATCTGAACCAAGGCTGTCAACGCG 1500
 Db 1441 AGCACTGAGTCTGTGCTTGTCACTGAAAGTTAGATCTGAACCAAGGCTGTCAACGCG 1500
 Qy 1501 TGTCTCAACTCCCACTCTGAGGAGATTTCCCTCATTTCTACCTCTCTAA 1560
 Db 1501 TGTCTCAACTCCCACTCTGAGGAGATTTCCCTCATTTCTACCTCTCTAA 1560
 Qy 1561 GGGTATGACCCCTCCCACTCTGAGGAGATTTCCCTCATTTCTACCTCTCTAA 1620
 Db 1561 GGGTATGACCCCTCCCACTCTGAGGAGATTTCCCTCATTTCTACCTCTCTAA 1620
 Qy 1621 CCCATCTCCCATCTGGAGATGAGGACCTTCATCAGCCTTAAACCTGGGAAATGCTGTG 1680
 Db 1621 CCCATCTCCCATCTGGAGATGAGGACCTTCATCAGCCTTAAACCTGGGAAATGCTGTG 1680
 Qy 1681 CCCCAGAGACTCTTGGTTTCTCCCAATACAGAGAGGAGGAGGAGGAGGAGGAGG 1740
 Db 1681 CCCCAGAGACTCTTGGTTTCTCCCAATACAGAGAGGAGGAGGAGGAGGAGGAGG 1740
 Qy 1741 GTCTAGTTAGAGGAGGCTCCAGGAGCAAGTCAAGCTCTCCCTCATGCTCTGTGTC 1800
 Db 1741 GTCTAGTTAGAGGAGGCTCCAGGAGCAAGTCAAGCTCTCCCTCATGCTCTGTGTC 1800

Db 1741 GTCTAGTTAGAGGAGGCTCCAGGAGCAAGTCAAGCTCTCCCTCATGCTCTGTGTC 1800
 Qy 1801 AGTGTGCTTAAAGGTGGCTCTCTACCTCCACACTGCTGGGCTCTGGGGAGAGACTGGGG 1860
 Db 1801 AGTGTGCTTAAAGGTGGCTCTCTACCTCCACACTGCTGGGCTCTGGGGAGAGACTGGGG 1860
 Qy 1861 AGGGGCGCTGGGAGAGCCCTGACGCTGGAACCTGTATACAAATAAAGACAGTCTAC 1920
 Db 1861 AGGGGCGCTGGGAGAGCCCTGACGCTGGAACCTGTATACAAATAAAGACAGTCTAC 1920
 Qy 1921 AGACNTCTGAGAGGCGCTGCTCCNGAGAGTCTCAACTTTAGGGGAGGAGGAGGAGGAGG 1980
 Db 1921 AGACNTCTGAGAGGCGCTGCTCCNGAGAGTCTCAACTTTAGGGGAGGAGGAGGAGGAGG 1980
 Qy 1981 AGAGAAA 1987
 Db 1981 AGAGAAA 1987
 RESULT 3
 ID ADF6734 standard; cDNA; 1987 BP.
 XX ADF6734;
 XX AC
 XX ADP6734;
 DT 12-FEB-2004 (first entry)
 XX
 DE cDNA encoding novel human protein NOV4.
 XX
 KW cytostratic; hepatotropic; vulnery; antipsoiatric; osteopathic;
 KW antiatherosclerotic; antiatherosclerotic; haemostatic; vasotropic;
 KW thrombolytic; antidiabetic; hypotensive; dermatological;
 KW immunosuppressive; antinephritic; immunosuppressive; fungicide;
 KW vincide; protozoicide; neuroprotective; antinephritic; antinephritic;
 KW antiatherosclerotic; antiparkinsonian; nootropic; anticonvulsant;
 KW NOV4 modulator; cancer; hyperproliferative disease; cirrhosis; keloid;
 KW psoriasis; tissue hypertrophy; osteoarthritis;
 KW atherosclerotic plaque formation; haemorrhage; ischaemic disease;
 KW thrombosis; diabetes mellitus; hypertension; hypothyroidism;
 KW immune deficiency; severe combined immunodeficiency; SCID; infection;
 KW malaria; candidiasis; autoimmune disorder; connective tissue disease;
 KW multiple sclerosis; systemic lupus; erythematous; rheumatoid arthritis;
 KW autoimmune pulmonary inflammation; Guillain-Barre syndrome;
 KW autoimmune thyroiditis; insulin dependent diabetes mellitus;
 KW myasthenia gravis; graft- versus-host disease;
 KW autoimmune inflammatory eye disease; asthma; haematopoiesis;
 KW tissue regeneration; wound healing; tissue repair; burn; incision; ulcer;
 KW periodontal disease; Alzheimer's disease; Parkinson's disease;
 KW Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
 KW human; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN US2003199103-A1.
 XX
 PD 23-OCT-2003.
 XX
 PF 15-OCT-2001; 2001US-00977639.
 XX
 PR 03-JUN-1999; 99US-0137322P.
 PR 15-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Fernandes E, Herrman J, Vernet C;
 XX
 DR WPI; 2004-021196/02.
 DR P-PSDB; ADF6735.
 XX

PT Novel substantially NOX polypeptide useful for diagnosing, preventing
PT and treating diseases e.g., cancer, multiple sclerosis, systemic lupus
erythematosus.

XX
XX
PS Disclosure; SEQ ID NO 7; 165pp; English.

CC The invention describes a substantially purified polypeptide (I) having
CC amino acid sequence chosen from a fully defined NOX sequence (S1) of 708
CC amino acids as given in the specification, or polypeptide having one or
CC more conservative amino acid substitutions of (S1), or mutant or variant
CC of (S1). (I) having (S1) is useful for diagnosing a pathological
CC condition associated with (I) or its activity in a subject e.g. cancer.
CC (I) useful in treatment of cancer, hyperproliferative diseases,
CC cirrhosis, keloid, psoriasis, tissue hypertrophy, osteoarthritis,
CC atherosclerotic plaque formation, haemorrhage, ischaemic heart or renal
CC disease, thrombosis, diabetes mellitus, hypertension, hypothyroidism. (I)
CC is useful in treatment of various immune deficiencies and disorders such
CC as severe combined immunodeficiency (SCID), bacterial infection, viral
CC infection such as herpes viral infection, protozoan infection such as
CC malaria, fungal infection such as candidiasis. (I) is also useful in
CC treating autoimmune disorders such as connective tissue disease, multiple
CC sclerosis, systemic lupus, erythematosus, rheumatoid arthritis,
CC autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune
CC thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis,
CC graft-versus-host disease and autoimmune inflammatory eye disease and
CC asthma. (I) useful in regulation of haematopoiesis, regeneration and
CC tissue growth of bone, cartilage, tendon, ligament and useful for wound
CC healing and tissue repair. (I) is also useful in treatment of burns,
CC incisions and ulcers. (I) also useful in treatment of periodontal
CC disease, Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, and Shy-Drager syndrome. (I) has effective
CC antitumor and anti-inflammatory activity. This sequence encodes a novel
CC human NOX protein.

XX
SQ Sequence 1987 BP; 448 A; 613 C; 532 G; 386 T; 0 U; 8 Other;

Query Match 99.6%; Score 1979; DB 12; Length 1987;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1987; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAATGAAGCGCTTCATGCTGTCAAGAGCCCATCCCAACCAAGCCCGGTGGGCA 60
Db 1 TGAATGAAGCGCTTCATGCTGTCAAGAGCCCATCCCAACCAAGCCCGGTGGGCA 60

QY 61 GGCACATCAAGACCTTAGAGAGCGCTATGAGTTGCGGTGAGCGTAGAGACTTCTCA 120
Db 61 GGCACATCAAGACCTTAGAGAGCGCTATGAGTTGCGGTGAGCGTAGAGACTTCTCA 120

QY 121 CTTGAAGCATCATTTGTGACCACTCCCAACCAACCAACATCGAGTGGCGGTGAGAA 180
Db 121 CTTGAAGCATCATTTGTGACCACTCCCAACCAACCAACATCGAGTGGCGGTGAGAA 180

QY 121 AGCTGGGGCTGAGAGGACATGTAATGAACCTTACGATACAGTGCAGTGCAGACT 240
Db 181 AGCTGGGGCTGAGAGGACATGTAATGAACCTTACGATACAGTGCAGTGCAGACT 240

QY 241 GCCGAGAGCGTGAACCCGAGCGTGGCGGTGACCTGGGCTACTGTGCGGAGAGAGCAG 300
Db 241 GCCGAGAGCGTGAACCCGAGCGTGGCGGTGACCTGGGCTACTGTGCGGAGAGAGCAG 300

QY 301 CCTACATAGTCCGGGACGGGTGACCCGCAATACAGAAACGTCAGAGACCTTCCGGAC 360
Db 301 CCTACATAGTCCGGGACGGGTGACCCGCAATACAGAAACGTCAGAGACCTTCCGGAC 360

QY 361 GGAGATCAAAATCTGAGTGCCTCTCCCTTCCCTTCCCTGTCCCGCCGCAAGCCTGC 420
Db 361 GGAGATCAAAATCTGAGTGCCTCTCCCTTCCCTTCCCTGTCCCGCCGCAAGCCTGC 420

QY 421 CAGCAAGCTCGCTAATCCCATTAACAAGCTCCAGACATCTCAGCCAGGTTCTAGC 480
Db 421 CAGCAAGCTCGCTAATCCCATTAACAAGCTCCAGACATCTCAGCCAGGTTCTAGC 480

QY 481 CCCACGACCCCGAGACCCCGAGTGAACATCTCCAAACTAGGCGCTTCACTATC 540

Db 481 CCCACGACCCCGAGACCCCGAGTGAACATCTCCAAACTAGGCGCTTCACTATC 540

QY 541 CAGGCGAGCGCAGGAGCTCCCTGGCTGACATGATGCCAGATTTGAGATTTGGCCTC 600
Db 541 CAGGCGAGCGCAGGAGCTCCCTGGCTGACATGATGCCAGATTTGAGATTTGGCCTC 600

QY 601 CTTCTATTATCCAGATTAACAGGCGCTGGGTCAGGAGAGAAATCTTAAGAACCACT 660
Db 601 CTTCTATTATCCAGATTAACAGGCGCTGGGTCAGGAGAGAAATCTTAAGAACCACT 660

QY 661 GTGGGTCAAGGGAATGGGACACGACGATATAGGGCAAGCTCTGACAGACAGACACA 720
Db 661 GTGGGTCAAGGGAATGGGACACGACGATATAGGGCAAGCTCTGACAGACAGACACA 720

QY 721 GACAAACCTCTGATCTATGAAATCTCTGACAGGCGCAAGGACCAAGGACCTGAACT 780
Db 721 GACAAACCTCTGATCTATGAAATCTCTGACAGGCGCAAGGACCAAGGACCTGAACT 780

QY 781 CTTGGCAAGGGAGTGGAGACAGAGGAGGTCAAGCAAGGAGTGGCTATCTAG 840
Db 781 CTTGGCAAGGGAGTGGAGACAGAGGAGGTCAAGCAAGGAGTGGCTATCTAG 840

QY 841 TGAATCTAATTCGCGAGGCTCAGCAAGGCGCAAGAGACAGCCGTGACGGTAACTT 900
Db 841 TGAATCTAATTCGCGAGGCTCAGCAAGGCGCAAGAGACAGCCGTGACGGTAACTT 900

QY 901 CCCCTTACAGGCTTCAAGCCCAAGGAGGAGAGGCTGCTGCCACCCCTGCCC 960
Db 901 CCCCTTACAGGCTTCAAGCCCAAGGAGGAGAGGCTGCTGCCACCCCTGCCC 960

QY 961 CCAGCAGCTGCTGACAGGCGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020
Db 961 CCAGCAGCTGCTGACAGGCGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020

QY 1021 CCAATACAGCTGTTGTGAAAATCTGATGAACTCTGCGCTGCGCTGCTGCGG 1080
Db 1021 CCAATACAGCTGTTGTGAAAATCTGATGAACTCTGCGCTGCGCTGCTGCGG 1080

QY 1081 CCTCAGGCGAAGGCGAAGTGGGCTGGGAGTGGGCTGGTCTTCCCTCCCAAGGCC 1140
Db 1081 CCTCAGGCGAAGGCGAAGTGGGCTGGGAGTGGGCTGGTCTTCCCTCCCAAGGCC 1140

QY 1141 TGTGTTTNTGGGCTGCTCCATGACAGAGATCACTTACAGATGAAAGCAAGGCC 1200
Db 1141 TGTGTTTNTGGGCTGCTCCATGACAGAGATCACTTACAGATGAAAGCAAGGCC 1200

QY 1201 ATGATGGGGCTTTGGGCTCTGAGGTTGAGACCCAGCTTCTTGCACTTCCCTCGG 1260
Db 1201 ATGATGGGGCTTTGGGCTCTGAGGTTGAGACCCAGCTTCTTGCACTTCCCTCGG 1260

QY 1261 CAGTACGCTCTCCATCCATCCCTCTTTAATCTATGAACTATAGGCTGGGTGTGTA 1320
Db 1261 CAGTACGCTCTCCATCCATCCCTCTTTAATCTATGAACTATAGGCTGGGTGTGTA 1320

QY 1321 ACAACACACCCCTATGCTGCTTCAATATCTACAGATTTACATTTGTTGAGGCAAT 1380
Db 1321 ACAACACACCCCTATGCTGCTTCAATATCTACAGATTTACATTTGTTGAGGCAAT 1380

QY 1381 TGAAGCTTTTCAATGAGATTTACATTTTCAATTTTCAATTTTCAATTTTCAATTTT 1440
Db 1381 TGAAGCTTTTCAATGAGATTTACATTTTCAATTTTCAATTTTCAATTTTCAATTTT 1440

QY 1441 AGCACTGAGTCTGCTGCTTTTCACTGAGGTTTGAATCTGAACCAAGGCTGCAACGC 1500
Db 1441 AGCACTGAGTCTGCTGCTTTTCACTGAGGTTTGAATCTGAACCAAGGCTGCAACGC 1500

QY 1501 TGTCTCACTCCCACTGAGGAGTGAAGATTTTCCCTCAATTTCTACTCTCTA 1560
Db 1501 TGTCTCACTCCCACTGAGGAGTGAAGATTTTCCCTCAATTTCTACTCTCTA 1560

QY 1561 GGCTATGACCCCTCCCACTTTTCACTGGGAGTGGGAGTCAATGAGAAAGCC 1620

Db 1561 GGGTATGACACCCCTCCACGCTTCCAGCTGGGGGATGGGGAGTCATAGGAAAAGCC 1620

Qy 1621 CCCATCTCCCATCTGGGATAGGAGACCTTCATCAGCCTTAACCTGGGAAATGCCCTGTG 1680

Db 1621 CCCATCTCCCATCTGGGATAGGAGACCTTCATCAGCCTTAACCTGGGAAATGCCCTGTG 1680

Qy 1681 CCCCCAGTACTCTTGTCTTCTCTCCCATCAGAGAGAGGGTGGAGGGAGGGTGG 1740

Db 1681 CCCCCAGTACTCTTGTCTTCTCTCCCATCAGAGAGAGGGTGGAGGGAGGGTGG 1740

Qy 1741 GTCTCAGTTAGCAGGGGGTCCCGAGGGCAAGTCACTCCCTCCCTCATGCTCTGTGGTC 1800

Db 1741 GTCTCAGTTAGCAGGGGGTCCCGAGGGCAAGTCACTCCCTCCCTCATGCTCTGTGGTC 1800

Qy 1801 AGTGTGCTTAAAGGGTGGGCTCTTCTCAGTCCCACTCTGGGCGCTTGGGGAGAGTGGGG 1860

Db 1801 AGTGTGCTTAAAGGGTGGGCTCTTCTCAGTCCCACTCTGGGCGCTTGGGGAGAGTGGGG 1860

Qy 1861 AGGGGGCGCTGGAGAGCCCTGACGCTGGAGCCTGATACAAATTAAGAGAGTCTAC 1920

Db 1861 AGGGGGCGCTGGAGAGCCCTGACGCTGGAGCCTGATACAAATTAAGAGAGTCTAC 1920

Qy 1921 AGACNTCTGGAGGCGCCCTGCGCNGAGTTCTCAACTTAAAGGCGAGGCGNNNCTTACTTG 1980

Db 1921 AGACNTCTGGAGGCGCCCTGCGCNGAGTTCTCAACTTAAAGGCGAGGCGNNNCTTACTTG 1980

Qy 1981 AGAGAAA 1987

Db 1981 AGAGAAA 1987

RESULT 4
ADO60244
ID ADO60244 standard; DNA; 1987 BP.

AC ADO60244;
XX
DT 15-JUL-2004 (First entry)
XX
DE Human NOV4 DNA.
XX
KW Human; NOV4 protein; cancer; hyperproliferative disease; cirrhosis;
KW keloid; psoriasis; tissue hypertrophy; osteoarthritis;
KW atherosclerotic plaque formation; haemorrhage; ischaemic heart disease;
KW renal disease; thrombosis; diabetes mellitus; hypertension;
KW hypothyroidism; severe combined immunodeficiency; SCID; infection;
KW malaria; candidiasis; autoimmune disorder; connective tissue disease;
KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW autoimmune pulmonary inflammation; Guillain-Barre syndrome;
KW autoimmune thyroiditis; myasthenia gravis; graft-versus-host disease;
KW autoimmune inflammatory eye disease; asthma; burn; incision; ulcer;
KW periodontal disease; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
KW haematopoiesis; wound healing; tissue repair; antitumor;
KW antiinflammatory; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH CDS 991..1449
FT /*tag= b
FT /product= "NOV4 protein"
FT /transl_except= (pos:1147..1149, aa:Xaa)
FT /note= "Xaa corresponds to any amino acid"
FT sig_peptide 991..1065
FT /*tag= a
FT /*tag= 1066..1446
FT /product= "Mature NOV4 protein"
XX
XX US2003134430-A1.
XX
XX 17-JUL-2003.

XX 15-OCT-2001; 2001US-00977751.
PF
XX 03-JUN-1999; 99US-0137322P.
XX
PR 16-MAR-2000; 2000US-0189810P.
XX
PR 22-MAR-2000; 2000US-0181158P.
PR 30-MAR-2000; 2000US-0191086P.
PR 03-MAY-2000; 2000US-0201388P.
PR 31-MAY-2000; 2000US-00584411.
XX
XX (CUBA-) CUBAGEN CORP.
XX
XX Shmkets RA, Fernandes E, Herrman J, Vernet C;
XX
XX WPI; 2004-068928/07.
XX
XX P-PSDB; ADO60245.
XX
XX Novel substantially purified NOV4 polypeptide for treating severe
XX combined immunodeficiency; candidiasis; cancer; asthma; multiple
XX sclerosis; systemic lupus erythematosus.
XX
XX Disclosure; SEQ ID NO 7; 155pp; English.
XX
XX
XX The invention relates to human NOV4 polypeptides and polynucleotides.
XX NOV4 sequences are useful in the treatment of cancer, hyperproliferative
XX diseases, cirrhosis, keloid, psoriasis, tissue hypertrophy,
XX osteoarthritis, atherosclerotic plaque formation, haemorrhage, ischaemic
XX heart or renal disease, thrombosis, diabetes mellitus, hypertension,
XX hypothyroidism, asthma, burns, incisions, ulcers, periodontal disease,
XX Alzheimer's disease, Parkinson's disease, Huntington's disease,
XX amyotrophic lateral sclerosis, Shy-Drager syndrome, immune deficiencies
XX and disorders such as severe combined immunodeficiency (SCID), bacterial
XX infection, viral infection e.g. herpes viral infection, protozoan
XX infection e.g. malaria, fungal infection e.g. candidiasis, autoimmune
XX disorders such as connective tissue disease, multiple sclerosis, systemic
XX lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
XX inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin
XX dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease
XX and autoimmune inflammatory eye disease. The invention is useful in
XX regeneration and tissue growth of bone, cartilage, tendon, ligament,
XX haematopoiesis regulation, wound healing and tissue repair. Sequences of
XX the invention also exhibit antitumor and antiinflammatory activities. The
XX present sequence is human NOV4 DNA.
XX
XX Sequence 1987 BP; 448 A; 613 C; 532 G; 366 T; 0 U; 8 Other;
XX
XX Query Match 99.6%; Score 1979; DB 12; Length 1987;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1987; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGAATGAAGGCTGTCCATGCTGTCGAAGAAGCCCATCCCAAGCCCGGTGGGA 60
Db 1 TGAATGAAGGCTGTCCATGCTGTCGAAGAAGCCCATCCCAAGCCCGGTGGGA 60
Qy 61 GGCACATCAAGACCTTAGAGAGCGCTTATGAGTTTGCAGTGAAGACTTTCTCA 120
Db 61 GGCACATCAAGACCTTAGAGAGCGCTTATGAGTTTGCAGTGAAGACTTTCTCA 120
Qy 121 CTTGAAGACATCATTTGTCAACACCTCCAAACCAACCAATGAGGTGGGGCTGAGA 180
Db 121 CTTGAAGACATCATTTGTCAACACCTCCAAACCAACCAATGAGGTGGGGCTGAGA 180
Qy 181 AGCTGCGGTGGAAGCAGACTGTATCATGAAACACCTTACGACTACACAAAGTGCAGACT 240
Db 181 AGCTGCGGTGGAAGCAGACTGTATCATGAAACACCTTACGACTACACAAAGTGCAGACT 240
Qy 241 GCGGAGAGACGTGAGACCCGAGCGTGGGTGACCTGCTACTGTGCGGAGGACGAGACG 300
Db 241 GCGGAGAGACGTGAGACCCGAGCGTGGGTGACCTGCTACTGTGCGGAGGACGAGACG 300
Qy 301 CCTCACTAGTCCGGGACGCGTACCCGATACAGAACAGTCCAGACAGACCTTCCGAGC 360
Db 301 CCTCACTAGTCCGGGACGCGTACCCGATACAGAACAGTCCAGACAGACCTTCCGAGC 360

Qy	361	GGAAATCAAATCTGAGGCTCTCCCTCTCCCTTCCCTGTCCGCCGCCCAAGCCTGC	420
Db	361	GGAGTCAAAATCTGAGTGGCTCTCCCTCTCCCTTCCCTGTCCGCCGCCCAAGCCTGC	420
Qy	421	CAGCAAAAGCTTCGTAAACCCCAATTACAAACAGTCCAGACATCTCAGCCCAAGTTCTAGC	480
Db	421	CAGCAAAAGCTTCGTAAACCCCAATTACAAACAGTCCAGACATCTCAGCCCAAGTTCTAGC	480
Qy	481	CCCCACGACCCCAAGACCCCAAGTGGACATCTCTCCCAAATAGAGGCCCTCCACTATC	540
Db	481	CCCCACGACCCCAAGACCCCAAGTGGACATCTCTCCCAAATAGAGGCCCTCCACTATC	540
Qy	541	CAGGCGAGGCCAGGGACCTCCCTGGCTTACACATGATGCCAGATTTAGATTGGCTTC	600
Db	541	CAGGCGAGGCCAGGGACCTCCCTGGCTTACACATGATGCCAGATTTAGATTGGCTTC	600
Qy	601	CGTCACTTAATCCAGAGTACAGAGGCTGGGGTTCAGGGAAAGAGATCTAAAGAAACCCACT	660
Db	601	CGTCACTTAATCCAGAGTACAGAGGCTGGGGTTCAGGGAAAGAGATCTAAAGAAACCCACT	660
Qy	661	GTGGGTCAAGGGGAATGGACCAAGCAGACATATGGGCAAGCTCTGCAGAGACAGACACA	720
Db	661	GTGGGTCAAGGGGAATGGACCAAGCAGACATATGGGCAAGCTCTGCAGAGACAGACACA	720
Qy	721	GACAAACCCCTTGATCTATGAAGTCTCTGCAGGGGCAAGGACCAAGGACCTGGAAACCT	780
Db	721	GACAAACCCCTTGATCTATGAAGTCTCTGCAGGGGCAAGGACCAAGGACCTGGAAACCT	780
Qy	781	CTTGGCCAGGGGAGTGGAGAGACAGAGGGAAGSTCAACAGGCAAGGCTGCTATCTAAG	840
Db	781	CTTGGCCAGGGGAGTGGAGAGAGACAGAGGGAAGSTCAACAGGCAAGGCTGCTATCTAAG	840
Qy	841	TGAAACTAATTGCCCCGAGGAGCTCAGCAAGGCCAAAGAGAGACAGCCGTGACGTAATACTT	900
Db	841	TGAAACTAATTGCCCCGAGGAGCTCAGCAAGGCCAAAGAGAGACAGCCGTGACGTAATACTT	900
Qy	901	CCCTCTTACAGCTCTCCAAAGCCCAAGCAGAGAGAGGCTGCTGCCACCCCTGTGCC	960
Db	901	CCCTCTTACAGCTCTCCAAAGCCCAAGCAGAGAGAGGCTGCTGCCACCCCTGTGCC	960
Qy	961	CCAGCCAGCTGAGCTGTGCGACAGGCGAGACCATGACATCTGTATATGATGGGGTTTT	1020
Db	961	CCAGCCAGCTGAGCTGTGCGACAGGCGAGACCATGACATCTGTATATGATGGGGTTTT	1020
Qy	1021	CCAATACAGCTGTGCTGTGAAAACTGATGAATCTCTGCGCTGCTGCTGTGAGG	1080
Db	1021	CCAATACAGCTGTGCTGTGAAAACTGATGAATCTCTGCGCTGCTGCTGTGAGG	1080
Qy	1081	CCTCAGGCAAGGCCAAGTGGGGGTGGGGGTGGGGTGGTCTTCTCCCTCCACAGGCC	1140
Db	1081	CCTCAGGCAAGGCCAAGTGGGGGTGGGGGTGGGGTGGTCTTCTCCCTCCACAGGCC	1140
Qy	1141	TGTGTTCTTGGGGCTGCTCCATGACAGAGATCACTTAACAGAGATGGAAGCCAGGC	1200
Db	1141	TGTGTTCTTGGGGCTGCTCCATGACAGAGATCACTTAACAGAGATGGAAGCCAGGC	1200
Qy	1201	ATGATAGGGGCTTGGGTCTCGAGAGTTGAAGCCAGCTTCTTGGCACTTCCCTCCGG	1260
Db	1201	ATGATAGGGGCTTGGGTCTCGAGAGTTGAAGCCAGCTTCTTGGCACTTCCCTCCGG	1260
Qy	1261	CAGTCAGCTTCCATCCATCCCTCTTTAATCTATGATCTATAGGCTCGGTGTGTGA	1320
Db	1261	CAGTCAGCTTCCATCCATCCCTCTTTAATCTATGATCTATAGGCTCGGTGTGTGA	1320
Qy	1321	ACAACACACCCCTATCGTGTCTTCAAAATCTCAGACTTACCAATTGGTTAGGCAAAAT	1380
Db	1321	ACAACACACCCCTATCGTGTCTTCAAAATCTCAGACTTACCAATTGGTTAGGCAAAAT	1380
Qy	1381	TCAGAGCTTTCTCAATCAGATTACATCTCCATTTTCAATTAACGAGGAAACATCCCGC	1440
Db	1381	TCAGAGCTTTCTCAATCAGATTACATCTCCATTTTCAATTAACGAGGAAACATCCCGC	1440

OY		1441	AGCCACTGAGTGCCTGTGGTTTGTCACGAAGGTTAGATCTGAACCCAGGGGTCAACNGC	1500
Db		1441	AGCCACTGAGTGCCTGTGGTTTGTCACGAAGGTTAGATCTGAACCCAGGGGTCAACNGC	1500
OY		1441	AGCCACTGAGTGCCTGTGGTTTGTCACGAAGGTTAGATCTGAACCCAGGGGTCAACNGC	1500
OY		1501	TGCTCTCAACTCCCCACCCTCTG9GGACA CTGAGAAGTATTCCCCTCATTTCTA CCTCTCTAA	1560
Db		1501	TGCTCTCAACTCCCCACCCTCTG9GGACA CTGAGAAGTATTCCCCTCATTTCTA CCTCTCTAA	1560
OY		1561	GCGTAGTACACCCCTCCGCCAGCTCTTCCAGTCTGGGGGATGGGGGAGTCA TGTGAAAAAGCC	1620
Db		1561	GCGTAGTACACCCCTCCGCCAGCTCTTCCAGTCTGGGGGATGGGGGAGTCA TGTGAAAAAGCC	1620
OY		1621	CCCATCTCCCACTGGGATAGGGACCTTCCATCAGCTTTAA CCTGGGAAAATGCTGCTG	1680
Db		1621	CCCATCTCCCACTGGGATAGGGACCTTCCATCAGCTTTAA CCTGGGAAAATGCTGCTG	1680
OY		1661	CCCCAGAGATCTTGTTTGTCTCCCACTAGAAAGAGGGGTGGAGGGGAGAGGGTGG	1740
Db		1661	CCCCAGAGATCTTGTTTGTCTCCCACTAGAAAGAGGGGTGGAGGGGAGAGGGTGG	1740
OY		1741	GTCCTAGTTTACAGAGGGGTCCCAAGGGCAAGTCAAGCTTCTCCATGCTCTCTG3TC	1800
Db		1741	GTCCTAGTTTACAGAGGGGTCCCAAGGGCAAGTCAAGCTTCTCCATGCTCTCTG3TC	1800
OY		1801	AGTGTGCTTTAGGGTGGCTTCTCACTCCACCACTCTGGGCTCTTG9GGAGGACTGGGG	1860
Db		1801	AGTGTGCTTTAGGGTGGCTTCTCACTCCACCACTCTGGGCTCTTG9GGAGGACTGGGG	1860
OY		1861	AGGGGGCCGTGGAGAGGCCCTGAGGCTGGAAACCTGTATCACATTAAGGACAGTCTCAC	1920
Db		1861	AGGGGGCCGTGGAGAGGCCCTGAGGCTGGAAACCTGTATCACATTAAGGACAGTCTCAC	1920
OY		1921	AGACNTCTGAGAGGCCGCTGCGCMGAGTCTCAAACCTTAGGGGCGAGGGCANNNTTACTTG	1980
Db		1921	AGACNTCTGAGAGGCCGCTGCGCMGAGTCTCAAACCTTAGGGGCGAGGGCANNNTTACTTG	1980
OY		1981	AGAGAAA 1987	
Db		1981	AGAGAAA 1987	
<hr/>				
RESULT 5				
AD119771				
ID	AD119771 standard; DNA; 1987 BP.			
XX AC	AD119771;			
XX DT	22-APR-2004 (first entry)			
DE	Human NOV4 DNA.			
XX KM	Secreted protein; NOX; diagnosis; metabolic disorder; diabetes; obesity;			
KM	infection; anorexia; cancer; cardiovascular disease; hypertension;			
KM	atherosclerosis; neurodegenerative disorder; Alzheimer's disease;			
KM	Parkinson's disease; epilepsy; immune disorder; osteoarthritis;			
KM	haemato-oncologic disorder; inflammatory skin disorder; asthma;			
KM	dyslipidemia; neurogenesis; cell differentiation; cell proliferation;			
KM	haematopoesis; wound healing; angiogenesis; chromosome mapping;			
KM	tissue typing; preventive medicine; pharmacogenomics; gene therapy;			
KM	anorectic; cardiant; virulence; antibacterial; fungicide; protozoacide;			
XX OS	nootropic; neuroprotective; dermatological; human; gene; ds.			
OS	Homo sapiens.			
FH	Key			
FT	Location/Qualifiers			
FT	CDS 991..1449			
FT	/*tag= b			
FT	/product= "Human NOV protein"			
FT	/transl_except= {pos:1171..11221, aa:Gly-Pro}			
FT	/transl_except= {pos:1144..1152, aa:Val-Gly}			
FT	sig_peptide 991..1152			
FT	/*tag= a			
FT	mat_peptide 1153..1446			

```

FT      /*tag= C
PN      /product= "Human mature NOV protein"
XX      US2004002134-A1.
XX      01-JAN-2004.
XX
PF      15-OCT-2001; 2001US-00977819.
XX
XX      03-JUN-1999; 99US-0137322P.
XX      16-MAR-2000; 2000US-0189810P.
XX      22-MAR-2000; 2000US-019158P.
XX      30-MAR-2000; 2000US-0193086P.
XX      03-MAY-2000; 2000US-0201388P.
XX      31-MAY-2000; 2000US-00584411.
XX
XX      (CURA-) CURAGEN CORP.
XX
PI      Shinkets RA, Fernandes ER, Herrman JL, Vernet CM;
XX
XX      WPI; 2004-070737/07.
XX      P-PSDB; ADI19772.
XX
PT      New NOVX nucleic acids encoding human KIAA0768 protein-like and human
PT      protein PRO-228 polypeptides, useful for treating NOVX-associated
PT      disorders.
XX
XX      Disclosure; SEQ ID NO 7; 95pp; English.
XX
CC      The present invention is based in part on the discovery of novel secreted
CC      and membrane-bound polypeptides and their encoding polynucleotides. The
CC      nucleic acids and polypeptides are collectively referred as NOVX. The
CC      invention is useful for treating, preventing and diagnosing diseases such
CC      as metabolic disorders, diabetes, obesity, infectious diseases such as
CC      viral, bacterial, fungal, helminthic and protozoal infections, anorexia,
CC      cancer, cardiovascular diseases such as hypertension and atherosclerosis,
CC      neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,
CC      epilepsy, immune disorders such as osteoarthritis, haematopoietic
CC      disorders, inflammatory skin disorders, asthma and various dyslipidemias.
CC      The invention is also useful as targets for the identification of small
CC      molecules that modulate or inhibit e.g. neurogenesis, cell
CC      differentiation, cell proliferation, haematopoiesis, wound healing and
CC      angiogenesis, as hybridisation probes, in chromosome mapping, tissue
CC      typing, preventive medicine and pharmacogenomics. The invention is also
CC      useful in gene therapy. The present sequence is human NOV DNA.
XX
SQ      Sequence 1987 BP; 448 A; 615 C; 530 G; 386 T; 0 U; 8 Other;
Query Match      99.4%; Score 1975.8; DB 12; Length 1987;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1985; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Db      301 CCTACTAGTCCGGGCAACGGGTCAACCCGCAATACAGAAACAGTCCAGAGACCTTCCGGAC 360
Qy      361 GGAGATCAAAATCTAGATGCTCTCCCTTCCCTTCCCTGTCCTCCCGCCACGCTGC 420
Db      361 GGAGATCAAAATCTAGATGCTCTCCCTTCCCTTCCCTGTCCTCCCGCCACGCTGC 420
Qy      421 CAGCAAGGCTCCGGTAAACCCATTAACAAGAGCTCCAGAGATCATCTACGCGCCAGGTTCCAG 480
Db      421 CAGCAAGGCTCCGGTAAACCCATTAACAAGAGCTCCAGAGATCATCTACGCGCCAGGTTCCAG 480
Qy      481 CCCACGCAACCCCAAGCCCAAGGTGACATCTCTCCCAATCTAGAGGCTCCATCTATC 540
Db      481 CCCACGCAACCCCAAGCCCAAGGTGACATCTCTCCCAATCTAGAGGCTCCATCTATC 540
Qy      541 CAGGGCAAGGCGCAGGAGCTCCCTGCTGACACATGATGCCAATTTCAATTTGGCTC 600
Db      541 CAGGGCAAGGCGCAGGAGCTCCCTGCTGACACATGATGCCAATTTCAATTTGGCTC 600
Qy      601 CGTCACTTAATCCAGATACAGAGGCGTGGGCTCAGAGAGAGAGATCTAAAGAACCACT 660
Db      601 CGTCACTTAATCCAGATACAGAGGCGTGGGCTCAGAGAGAGAGATCTAAAGAACCACT 660
Qy      661 GTGGGTACGGGGAATGGACACAGCAGACATATGGGCAAGCTTGCAGACACAGACAC 720
Db      661 GTGGGTACGGGGAATGGACACAGCAGACATATGGGCAAGCTTGCAGACACAGACAC 720
Qy      721 GACAAACCTCTGATCTATGATAGTCTGACAGGCAAGGGGACAGGACCTTGGAAACCT 780
Db      721 GACAAACCTCTGATCTATGATAGTCTGACAGGCAAGGGGACAGGACCTTGGAAACCT 780
Qy      781 CTTGGCCAAAGGGAGTGGGAGAGACAGAGGAAAGTTCACAGGCAAGGTTCTTATTAAG 840
Db      781 CTTGGCCAAAGGGAGTGGGAGAGACAGAGGAAAGTTCACAGGCAAGGTTCTTATTAAG 840
Qy      841 TGGAACTAATTTCCCGGAGGCTCAGACAGGCAAGGAGAGACAGCGGTGACGTTAACT 900
Db      841 TGGAACTAATTTCCCGGAGGCTCAGACAGGCAAGGAGAGACAGCGGTGACGTTAACT 900
Qy      901 CCCCTTACAGGCTCTCCAGACCCCAAGCAGGAGAGAGGCTGCTCCCAACCCCGTGC 960
Db      901 CCCCTTACAGGCTCTCCAGACCCCAAGCAGGAGAGAGGCTGCTCCCAACCCCGTGC 960
Qy      961 CCAGCCAGCTGGCTGTGTCACAGGAGAGGACCATCTGATATGATAGTGGGTTT 1020
Db      961 CCAGCCAGCTGGCTGTGTCACAGGAGAGGACCATCTGATATGATAGTGGGTTT 1020
Qy      1021 CCAATACAGCTGGTCTGAAAAAATGATGAATCTCTGCGCTGCTGCTGCGG 1080
Db      1021 CCAATACAGCTGGTCTGAAAAAATGATGAATCTCTGCGCTGCTGCTGCGG 1080
Qy      1081 CTTCCAGGCAAGGCAAGTGGGCTGGGCTGGTCTCTTCTCTCCCAAGGCT 1140
Db      1081 CTTCCAGGCAAGGCAAGTGGGCTGGGCTGGTCTCTTCTCTCCCAAGGCT 1140
Qy      1141 TGTGTTCTGTTGGGCTGCTCCATGACAGACAGATCTACCTTACAGATGGAACCGAGG 1200
Db      1141 TGTGTTCTGTTGGGCTGCTCCATGACAGACAGATCTACCTTACAGATGGAACCGAGG 1200
Qy      1201 ATGATGAGGCTTTGGGCTCTGAGGTTGAGACCCAGCTCTTCCACCTCCCTCCGG 1260
Db      1201 ATGATGAGGCTTTGGGCTCTGAGGTTGAGACCCAGCTCTTCCACCTCCCTCCGG 1260
Qy      1261 CAGTCAAGCTCTCCATCATCCCTCTTTAATCTATGAACTCTATAGCTGGTGTGTGTA 1320
Db      1261 CAGTCAAGCTCTCCATCATCCCTCTTTAATCTATGAACTCTATAGCTGGTGTGTGTA 1320
Qy      1321 ACAACACACCCCTATGCTTCTCTCAAAATCTACAGATTCATTCATTTGAGGCCAAT 1380
Db      1321 ACAACACACCCCTATGCTTCTCTCAAAATCTACAGATTCATTCATTTGAGGCCAAT 1380
Qy      1381 TCAGAGCTTCTCAATTCATTTCAATTCATTTCAATTTCAATTTCAATTTCAATTTCCCG 1440
Db      1381 TCAGAGCTTCTCAATTCATTTCAATTCATTTCAATTTCAATTTCAATTTCAATTTCCCG 1440

```


Db 181 AGTGCCTGACGACGACCTGTACATGAAACCTTACGACTACACAAAGTGCAGACT 240
QY 241 GCCGAGGAGCGTGAACCCCGAGCGTGCCTGCTACTGTGAGGAGACGAGACG 300
Db 241 GCCGAGGAGCGTGAACCCCGAGCGTGCCTGCTACTGTGAGGAGACGAGACG 300
QY 301 CTTCACTAGTCCGGGACCGGCTACCCCGCATACAGAACGTCACAGACCTTCCGAG 360
Db 301 CTTCACTAGTCCGGGACCGGCTACCCCGCATACAGAACGTCACAGACCTTCCGAG 360
QY 361 GGAGATCAAAATGTAGTGGCTCTCCCTTCCCTTCCCTGTCGCCCGCCGACGCTGC 420
Db 361 GGAGATCAAAATGTAGTGGCTCTCCCTTCCCTTCCCTGTCGCCCGCCGACGCTGC 420
QY 421 CAGCAAGCTCGCTTAACCCCATTTACACAGCTCCAGACATCTACGCCAGGTTCTAGC 480
Db 421 CAGCAAGCTCGCTTAACCCCATTTACACAGCTCCAGACATCTACGCCAGGTTCTAGC 480
QY 481 CCCGAGGACCCGAGACCCGAGTGAACCTCTCCCAACTAGGGGCTCCGACCTTATC 540
Db 481 CCCGAGGACCCGAGACCCGAGTGAACCTCTCCCAACTAGGGGCTCCGACCTTATC 540
QY 541 CAGGAGGACGACGAGGACTCCCTGCTGACACATGATGCCAGATTTGAGTTGCTC 600
Db 541 CAGGAGGACGACGAGGACTCCCTGCTGACACATGATGCCAGATTTGAGTTGCTC 600
QY 601 CGTCACTTATCCAGAGTACAGGAGCTGGGCTCAGGAGAGAAATCTTAAGACCACT 660
Db 601 CGTCACTTATCCAGAGTACAGGAGCTGGGCTCAGGAGAGAAATCTTAAGACCACT 660
QY 661 GTGGGTCAAGGAGTGGGACACAGACAGACATATGAGCAAGCTCTGACAGACAGACA 720
Db 661 GTGGGTCAAGGAGTGGGACACAGACAGACATATGAGCAAGCTCTGACAGACAGACA 720
QY 721 GACAAACCTCTCTATCTATGAAAGTCTCTGACAGGAGGAGGACCTGAGAACCT 780
Db 721 GACAAACCTCTCTATCTATGAAAGTCTCTGACAGGAGGAGGACCTGAGAACCT 780
QY 781 CTGGCCAGAGGAGTGGGAGACAGAGGAGGAGTCAAGGAGGAGGAGGAGGAGGAG 840
Db 781 CTGGCCAGAGGAGTGGGAGACAGAGGAGGAGTCAAGGAGGAGGAGGAGGAGGAG 840
QY 841 TGGAACTAATGCTCCGAGGCTCAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
Db 841 TGGAACTAATGCTCCGAGGCTCAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
QY 901 CCCCTTACAGGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
Db 901 CCCCTTACAGGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
QY 961 CGAGCAGGCTGCTGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
Db 961 CGAGCAGGCTGCTGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
QY 1021 CGAATACAGTGGTGTGTAAGAACTGCAATGAAACCTGCGCTGCTGCGCTGCGG 1080
Db 1021 CGAATACAGTGGTGTGTAAGAACTGCAATGAAACCTGCGCTGCTGCGCTGCGG 1080
QY 1081 CCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
Db 1081 CCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
QY 1141 TGTGTCNTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
Db 1141 TGTGTCNTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
QY 1201 ATGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
Db 1201 ATGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
QY 1261 CAGTCACTCTCCATCCATCCCTCTTTAATCTATGAATCTATGAGGAGGAGGAGGAG 1320
Db 1261 CAGTCACTCTCCATCCATCCCTCTTTAATCTATGAATCTATGAGGAGGAGGAGGAG 1320

QY 1321 ACAACACACCCCTATCGTGTCTCTCAAACTACAGATTACCATTTGAGGAGGAGGAG 1380
Db 1321 ACAACACACCCCTATCGTGTCTCTCAAACTACAGATTACCATTTGAGGAGGAGGAG 1380
QY 1381 TCAGAGCTTCTCAAACTACAGATTACCATTTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
Db 1381 TCAGAGCTTCTCAAACTACAGATTACCATTTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
QY 1441 AGCCACTGAGTGTGTCTTGTCTGATGAAAGTTAGATCGAACCCAGGAGGAGGAGGAG 1500
Db 1441 AGCCACTGAGTGTGTCTTGTCTGATGAAAGTTAGATCGAACCCAGGAGGAGGAGGAG 1500
QY 1501 TGTCTCAACTCCCGACCTCTGGGACACTGAGAGGATTTCCCTCATTTACTCTCTTAA 1560
Db 1501 TGTCTCAACTCCCGACCTCTGGGACACTGAGAGGATTTCCCTCATTTACTCTCTTAA 1560
QY 1561 GGCTATGACCCCTCCCGACGCTCTCCAGCTGGGAGGAGTGAAGGAGGAGGAGGAGGAG 1620
Db 1561 GGCTATGACCCCTCCCGACGCTCTCCAGCTGGGAGGAGTGAAGGAGGAGGAGGAGGAG 1620
QY 1621 CCCATCTCCCATCTGGGATGAGGACCTTCCATCAGCTTAACTCTGGGAGGAGGAGGAG 1680
Db 1621 CCCATCTCCCATCTGGGATGAGGACCTTCCATCAGCTTAACTCTGGGAGGAGGAGGAG 1680
QY 1681 CCCCAGTGACTCTTGTCTGCTCTCCCATACAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
Db 1681 CCCCAGTGACTCTTGTCTGCTCTCCCATACAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
QY 1741 GTCTCACTTACAGAGGAGTCCCGACAGGAGGAGTCAAGCTCTCTCTCTCTCTCTCT 1800
Db 1741 GTCTCACTTACAGAGGAGTCCCGACAGGAGGAGTCAAGCTCTCTCTCTCTCTCTCT 1800
QY 1801 AGTGTGCTTACAGGAGTGGGCTCTCACTCCACCACTCTGGGAGGAGGAGGAGGAGGAG 1860
Db 1801 AGTGTGCTTACAGGAGTGGGCTCTCACTCCACCACTCTGGGAGGAGGAGGAGGAGGAG 1860
QY 1861 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
Db 1861 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
QY 1921 AGAATCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
Db 1921 AGAATCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
QY 1981 AGAGAAA 1987
Db 1981 AGAGAAA 1987

RESULT 7
AAZ3583
ID AAZ3583 standard; cDNA; 2192 BP.
XX AAZ3583;
XX AC
XX AAZ3583;
DT 08-DEC-1999 (first entry)
XX
XX Human breast tumour-associated EST 43.
XX
XX Expressed sequence tag; EST; human; breast; cancer; cytostatic;
XX
XX medicaments; gene therapy; treatment; fat metabolism; ss.
OS Homo sapiens.
XX
XX DE19813835-A1.
XX
XX
XX 23-SEP-1999.
XX
XX
XX 20-MAR-1998; 98DE-01013835.
XX
XX
XX 20-MAR-1998; 98DE-01013835.
XX
XX

PA (META-) METAGEN GES GENOMFORSCHUNG MBH
XX

PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;

DR WPI; 1999-528979/45
DB P-BDDB; AYV49E01 A

DR P-PSDB; AAY48501, AAY48502.
yy

PT	Human nucleic acid sequences and protein products from normal breast
PT	tissue, useful for breast cancer therapy.
VV	

PS Claim 1a; 131; 206pp; German.

This invention describes novel human nucleic acid sequences from normal breast tissue which have cytostatic activity. The nucleic acid sequences can be used to produce and isolate full-length gene sequences. They can be used to express proteins, which can be used as tools to find an activity against breast cancer. The sequences can be used in sense or antisense form. They are especially useful for medicaments for gene therapy to treat breast cancer and for treating illnesses associated with metastasis. AA33541-733610 represent expressed sequence tags described in the method of the invention

Sequence 2192 BP; 457 A; 717 C; 588 G; 430 T; 0 U; 0 Other;

Query Match	88.8%	Score 1684.4	DB 2	Length 2192
Best Local Similarity	95.0%	Pred. No. 0		
Matches 1658; Conservative	0	Mismatches 73	Indels 24	Gaps 11

Qy	33	GCCCATCCCAACCAACCCCGGGGCAAGCAACATCAAGACCTTAGAGACGCTATGA	92
Db	258	GGCCTTCCAGCCCGCCGGTGGGCAAGCAACATCAAGACCTTAGAGACGCTATGA	317
Qy	93	GTTTGGGTGGAGGTGAGAGACTCTCACCCTGAAGACATCATTTGTACCACTCTCAACA	155
Db	318	GTTTGGGTGGAGGTGAGAGACTCTCACCCTGAAGACATCATTTGTACCACTCTCA--A	374
Qy	153	AACCAACATCAGAGTGGGGGCTGAGAAAGCTGGGGCTGAACGACACTGTGACATGAAA	212
Db	375	CAACCAACATCAGAGTGGGGG--CTGAGAGCTGGCGGCTGAACGACACGTCATGAACCC	433
Qy	213	CACCTTACGACTACAGATGGCCAGCTACCGAGAGAGGTGGACCCCGACGGTCCGGTGA	272
Db	433	TTGGCT-----CAAGATGCCAG--CTCCCGAGAGACGTGAACCC--GACGTGGTGA	487
Qy	273	CTTGGCTACTGTGGGAGGAGCGAGCAGCTCTCACTAGTCCGGGACGCGTCACCCGATA	332
Db	482	CTTCGG--CTTGGCGGAGAGAGCGAGAGCTCACTACCGGCGACGGGTGACCCGATA	538
Qy	333	CAGAACACGTCCAGCAGACCTTCCGAGCGAGTCAAAATCTGATGCTCTCCCTTCCC	392
Db	539	CAGAACACGTCCAGCAGACCTTCCGAGCGAGTCAAAATCTGATGCTCTCCCTTCCC	598
Qy	393	TTTTCCTCT--CCCCCGCCCCACAGCTGCTCCAGCAAGAAAGCTGCTAACCCATTCAACAG	451
Db	599	TTTTCCTCTGCCCCCGCCCCACAGCTGCTCCAGCAAGAAAGCTGCTAACCCATTCAACAG	658
Qy	452	CTCCAGACATCTCAGCCAGGTTCTAGCCCCCAGCAGCACCAGACCCCAAGTGAACAT	511
Db	659	CTCCAGACATCTCAGCCAGGTTCTAGCCCCCAGCAGCACCAGACCCCAAGTGAACAT	718
Qy	512	CTTCCCAAATATAGGGCTTCAATCTATCAGGCGAGGCGAGGACTCTCCCTGGCTGACA	572
Db	719	CTTCCCAAATATAGGGCTTCAATCTATCAGGCGAGGCGAGGACTCTCCCTGGCTGACA	778
Qy	572	CATGATCCCCAGATTTCAGATTGGCTCCGTCATCTATTCAGAGTACAGGGGCTGGGG	631
Db	779	CATGATCCCCAGATTTCAGATTGGCTCCGTCATCTATTCAGAGTACAGGGGCTGGGG	838
Qy	632	TCAGGGAAGGAAGTCTAAAGAACCACTGTGGGTAGGGGAATGGGACACAGCAGACAT	692
Db	839	TCAGGGAAGGAAGTCTAAAGAACCACTGTGGGTAGGGGAATGGGACACAGCAGACAT	899
Qy	692	ATGGGCAAGCTCTTGACGAGCAGACAGACAAACCTCTGATCTTAAAGTCTCTGCA	751

Db	899	ATGGGCAAGCTCTGCAAGACAGACAGGCAAGCAAAACCTCTGATCTCTATGAAGTCTTGCA	958
OY	752	GGGCAAGGGACCAAGGAGACCTTGAAACCTCTTGGCCAAAGGGAAATGGAGAGACAGAGGG	811
Db	959	GGGCAAGGGGACCAAGGAGACCTTGAAACCTCTTGGCCAAAGGGAAATGGAGAGAGAGGG	1018
OY	812	AAGGTCAAGGCAAGGGTGGCTTATCTAATGTAAGTAATATTGGCCAGAGGGCTCAGCAAGGC	871
Db	1019	AAGGTCAAGGCAAGGGTGGCTTATCTAATGTAAGTAATATTGGCCAGAGGGCTCAGCAAGGC	1078
OY	872	CAAGAGAGACAGCGCTGAGCGGTAACTTCCCTCTACAGCTTCAAGCCCAAGCCAGCCAG	931
Db	1079	CAAGAGAGACAGCGCTGAGCGGTAACTTCCCTCTACAGCTTCAAGCCCAAGCCCAAGCCAG	1138
OY	932	CGACAGAGCTGCGCCCAAGCCCGGTGGCCCAAGCACTGAGTGGTGGCAGGGCAGAGCCA	991
Db	1139	CGACAGAGCTGCGCCCAAGCCCGGTGGCCCAAGCACTGAGTGGTGGCAGGGCAGAGCCA	1198
OY	992	TGCCACATCTGTATATAGATGGGGTTTTTCCAAATACAGCTGGTTGTAAGAAAATCTGCATG	1051
Db	1199	TGCCACATCTGTATATAGATGGGGTTTTTCCAAATACAGCTGGTTGTAAGAAAATCTGCATG	1258
OY	1052	AAATCTCTGCGCTCTGCGCTGCTGGGGGCTCCAGGCAAGGCCAAGTGGGCTTGGGGGT	1111
Db	1259	AAATCTCTGCGCTCTGCGCTGCTGGGGGCTCCAGGCAAGGCCAAGTGGGCTTGGGGGT	1318
OY	1112	GGGCGCTGGTCTCTTCCCTCCCAAGGCGCTGTCTTCTTCTGAGGGGCTGCTCCATGACAGAG	1171
Db	1319	GGGCGCTGGTCTCTTCCCTCCCAAGGCGCTGTCTTCTTCTGAGGGGCTGCTCCATGACAGAG	1378
OY	1172	GATCACTTACAGAGATGGAAGCCAGGCGATGATGGGGCTTTGGGTCCTGAGAGTTGGA	1231
Db	1379	GATCACTTACAGAGATGGAAGCCAGGCGATGATGGGGCTTTGGGTCCTGAGAGTTGGA	1438
OY	1232	CCCCAGCTTCTGGCACCTTCCCCCTCC-GGGAGAGCAGTCTCCAAATCAGCCCTCTTTA	1290
Db	1439	CCCCAGCTTCTGGCACCTTCCCCCTCCGGGAGAGCAGTCTCCAAATCAGCCCTCTTTA	1498
OY	1291	ATCTATGATCTATAGAGCTGGTGTGTGAACA-ACAACCCCAATCGTGTCTTCAAA	1349
Db	1499	ATCTATGATCTATAGAGCTGGTGTGTGAACAACAACAACCCCAATCGTGTCTTCAAA	1558
OY	1350	TATCAGACATTAACATTTGGTTGAGGCCAAATTCAGACCTTCTCAAAATCAGATTTCAGAT	1409
Db	1559	TATCAGACATTAACATTTGGTTGAGGCCAAATTCAGACCTTCTCAAAATCAGATTTCAGAT	1618
OY	1410	CTCCATTTTCTTAACGGGGAAACATCCCGCAGCAGCTGAGTGTGTGCTTTGTCACATGA	1469
Db	1619	CTCCATTTTCTTAACGGGGAAACATCCCGAGCAGCTGAGTGTGTGCTTTGTCACATGA	1678
OY	1470	AGGTTAATCTGAACCCAGGGGTGTCAAACNGTCTCTCAGCTCCCACTCTTGGGCACTG	1529
Db	1679	AGGTTAATCTGAACCCAGGGGTGTCAAACAGTGTCTCAGCTCCCACTCTTGGGCACTG	1738
OY	1530	AGGAGTATTTTCCCTCAATCTAAGCTCTCTAAGGCTTAAGACCCCTCCCAAGTCTTCAG	1589
Db	1739	AGGAGTATTTTCCCTCAATCTAAGCTCTCTAAGGCTTAAGACCCCTCCCAAGTCTTCAG	1798
OY	1590	CTGGGGGATGGGGGAGTCAATAGAAAAACCCCATCTCCCATCTGGGATAGGAGACTTC	1649
Db	1799	CTGGGGGATGGGGGAGTCAATAGAAAAACCCCATCTCCCATCTGGGATAGGAGACTTC	1858
OY	1650	CATACAGCTTAACCTCTGGGAAATGCGTGGCCCAAGTACATCTTGGTTTCGTCTCCA	1709
Db	1859	CATACAGCTTAACCTCTGGGAAATGCGTGGCCCAAGTACATCTTGGTTTCGTCTCCA	1918
OY	1710	CATACAGAGAGGAGTGAAGGGGAAAGGTTGGTCTCAGTTAGCAAGGGATCCCAAGGCCAA	1769
Db	1919	CATACAGAGAGGAGTGAAGGGGAAAGGTTGGTCTCAGTTAGCAAGGGATCCCAAGGCCAA	1978
OY	1770	GTCAGCTCTCCCTCCATGCTCTCTGTGACAGTGTGCTTAAGGTTGGCTTCACTCCC	1829

Db 1979 GTACGCTCTCCCTCCATGCTCTGTGTCAGTGTGCTTAGGATGAGGCTCTCACTCCC 2038
Qy 1830 ACCACTTTGGG-CCCTTGGGGGAGACTGGGGAGGGGGCCCTGGGAGAGCCCTGAGGCTG 1888
Db 2039 ACCACTTGGAGCCCTTGGGGAGAGACTGGGGAGGGGGCCCTGGGAGAGCCCTGAGGCTG 2098
Qy 1889 GAACTGTATATACATATMAAGAGACAGTCTCACAGACNTCTGAGGCGCGCTGCGGAGT 1948
Db 2099 GAACCTGTATATACATATMAAGAGACAGTCTCACAGACMAAAGAGCGCGCTGCG- 2157
Qy 1949 TCTCAAACTTAGGGCAGGCGGNNNNCTTACTTGAGA 1983
Db 2158 TCTCAAACTTAGGGCAGGCGCTTACTTGAGAGAAA 2192

RESULT 8
AA160565
ID AA160565 standard; cDNA; 2142 BP.
XX
AC AA160565;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4554.
XX
XX Human, neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemocarcin;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000MO-US034263.
XX
XX 23-DEC-1999; 99US-00471275.
XX 21-JAN-2000; 2000US-00488725.
XX 25-APR-2000; 2000US-00552317.
XX 20-JUN-2000; 2000US-00598042.
XX 19-JUL-2000; 2000US-00620312.
XX 03-AUG-2000; 2000US-00653450.
XX 14-SEP-2000; 2000US-00662191.
XX 19-OCT-2000; 2000US-00693036.
XX 29-NOV-2000; 2000US-00727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
XX Wang Z, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA,
XX Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
XX P-PSDB; AAM41409.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
XX
XX Claim 1; SEQ ID NO 4554; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and the
XX encoded polypeptides (AAM38642-AA42213) with neotropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 2142 BP; 451 A; 700 C; 568 G; 423 T; 0 U; 0 Other;
Query Match 84.0%; Score 1670; DB 4; Length 2142;
Best Local Similarity 95.8%; Pred. No. 0;
Matches 1815; Conservative 0; Mismatches 57; Indels 22; Gaps 9;
Qy 33 GCCATCCCAACCAAGCCCGGTGGGAGGAGCAATCAAGACCTTAGAGAGCGCTTGA 92
Db 259 GGCCTCCCAAGCCCGCCCGGTGGGAGGAGCAATCAAGACCTTAGAGAGCGCTTGA 318
Qy 93 GTTTCGGGTGAGCTGAGAGACTTCTCACTGAAACATATTGTCACTCACTTCAACA 152
Db 319 GTTTCGGGTGAGCTGAGAGACTTCTCACTGAAACATATTGTCACTCACTTCAACA 375
Qy 153 AACCAACATGAGGTGGGGGCTGAGAAAGTGGCGGCTGAGAGGACATGTACATGAAA 212
Db 376 CAACCATGAGGTGGGG-CTGAGAAAGTGGCGGCTGAGAGGACATGTACATGAAAC 433
Qy 213 CACCTTACGACTACAAAGTCCAGACTGCGGAGAGAGCTGAGACCCCGAGCGGTGA 272
Db 434 TTCCCT-----CACAAGTCCAG-CTGCCGAGAGACGTGAGACC--GACGTGAGTGA 482
Qy 273 CCTGCGGCTAGTGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 332
Db 483 CTTGG--CTTGGGAG 539
Qy 333 CAGAACAGTCCAGACAGACCTTCGAGAGAGATCAAAATCTAGTGCCTTCTCC 392
Db 540 CAGAACAGTCCAGACAGACCTTCGAGAGAGATCAAAATCTAGTGCCTTCTCC 599
Qy 393 TTTCCCTGTCCCCCGCCCGCCAGAGCTGCGCAAGAGCTGCTAACCCCAATTAACAAG 452
Db 600 TTTCCCTGTCCCCCGCCCGCCAGAGCTGCGCAAGAGCTGCTAACCCCAATTAACAAG 659
Qy 453 TCCAGACATCTCAGCCAGAGCTTCTAGCCCGCCAGAGACCCAGAGAGAGAGAGAG 512
Db 660 TCCAGACATCTCAGCCAGAGCTTCTAGCCCGCCAGAGACCCAGAGAGAGAGAGAG 719
Qy 513 CTCCCAACTAGAGGCGCTTCACTCTATCCAGGAGAGAGAGAGAGAGAGAGAGAGAG 572
Db 720 CTCCCAACTAGAGGCGCTTCACTCTATCCAGGAGAGAGAGAGAGAGAGAGAGAGAG 779
Qy 573 ATGATGCCAGATTTCAGATTTCGAGCTTCGCTCACTTAATCAGAGTACAGGGGCTGAG 632
Db 780 ATGATGCCAGATTTCAGATTTCGAGCTTCGCTCACTTAATCAGAGTACAGGGGCTGAG 839
Qy 633 CAGGAGAGAGAGATCTTAAGAACCCAGTGTGAGGAGAGAGAGAGAGAGAGAGAGAG 692
Db 840 CAGGAGAGAGAGATCTTAAGAACCCAGTGTGAGGAGAGAGAGAGAGAGAGAGAGAG 899
Qy 693 TGGGCAAGCTCTGAG 752
Db 900 TGGGCAAGCTCTGAG 959
Qy 753 GGCAGAGGAGACAG 812
Db 960 GGCAGAGGAGACAG 1019
Qy 813 AGGTACAG 872
Db 1020 AGGTACAG 1079
Qy 873 AAG 932
Db 1080 AAG 1139

QY	933	GAGCAGGCTGCTCCACCCCGGCGCCCGCCAGCAGCTGTGGCCAGGGCAAGCCAT	992
Db	1140	GAGCAGGCTGCTCCACCCCGGCGCCCGCCAGCAGCTGTGTGCCAGGGCAAGCCAT	1139
QY	993	GCCACATCTGTATATAGATGGGGTTTTTCCAAATACAGTGGTGTGTGAAAACATGCATGA	1052
Db	1200	GCCACATCTGTATATAGATGGGGTTTTTCCAAATACAGTGGTGTGTGATTAACATGCATGA	1259
QY	1053	AACCTCGCGGCTCCTGGGCGCTGTGGGCGCTCCAGGCAAGGCAAGTGGGGTTTGGGGGTG	1112
Db	1260	AACCTCTGCGCTCCTGGGCGCTGTGGGCGCTCCAGGCAAGGCAAGTGGGGTTTGGGGGTG	1319
QY	1113	GGGCTGGTCCCTCTCCCTCCCAAGGCGCTGTGTCTMTGGGGCTGTCTCCATGCAAGACGG	1172
Db	1320	GGGCTGGTCCCTCTCCCTCCCAAGGCGCTGTGTCTTGGGGCTGTCTCCATGCAAGACGG	1379
QY	1173	ATCACCTTAACAAGATGGAAAGCCAGGGGATGGATGGGGCTTTGGGTCTCTGAGGTGGAC	1232
Db	1380	ATCACCTTAACAAGATGGAAAGCCAGGGGATGGATGGGGCTTTGGGTCTCTGAGGTGGAC	1439
QY	1233	CCGAGCTTCTTCCGACCTTCCCTCC-GGCAGTAGCTCTCCATCATCCCTCTTTAA	1291
Db	1440	CCGAGCTTCTTCCGACCTTCCCTCCCGGCGAGTAGCTCTCATCATCCCTCTTTAA	1439
QY	1292	TCATAGATCTATAGGCTCGGTGTGTGAACA-AACAACCCCTATCGTTGTCTTCAAT	1356
Db	1500	TCATAGATCTATAGGCTCGGTGTGTGAACAACAACCCCTATCGTTGTCTTCAAT	1559
QY	1351	ACTAGACATTAACAATTGGTTGAGGCAAAATCAGAGCTTCTTCAATAGATTAACAATC	1410
Db	1560	ACTAGACATTAACAATTGGTTGAGGCAAAATCAGAGCTTCTTCAATAGATTAACAATC	1619
QY	1411	TCGATTTTCATTAACGGGGAAACAATCCCGGACCACTGAGTGTGTCTTTGTCACTGAA	1470
Db	1620	TCGATTTTCATTAACGGGGAAACAATCCCGGACCACTGAGTGTGTCTTTGTCACTGAA	1679
QY	1471	GGTTAGATCTGAACCCAGGGGTGCAACNGCTGTCTGAATCTCCCACTCTGGGACACTGA	1530
Db	1680	GGTTAGATCTGAACCCAGGGGTGCAACAGCTGTCTGAATCTCCCACTCTGGGACACTGA	1739
QY	1531	GGAGTATTTCCCTCATCTCACTCTCTAAGGCTATGCAACCCCTCCACGCTTCCAGC	1580
Db	1740	GGAGTATTTCCCTCATCTCTAAGGCTATGCAACCCCTCCACGCTTCCAGC	1739
QY	1591	TGGGGGATGGGGGAGTCATAGGAAAGCCCCCATCTCCCATCTGGGATAGGGACCTTCC	1650
Db	1800	TGGGGGATGGGGGAGTCATAGGAAAGCCCCCATCTCCCATCTGGGATAGGGACCTTCC	1858
QY	1651	ATCAGCCCTTAACCCCTGGGAAATGCGCTGTGCCCCAGAGCACTTGTGTGTCTCCAC	1710
Db	1860	ATCAGCCCTTAACCCCTGGGAAATGCGCTGTGCCCCAGAGCACTTGTGTGTCTCTCCAC	1919
QY	1711	ATACAGAAGCAGGCTGAGGAGGAAAGGTGGGTCTCAGTTAGCAGGGGTCCCCAGGGCAAG	1770
Db	1920	ATACAGAAGCAGGCTGAGGAGGAAAGGTGGGTCTCAGTTAGCAGGGGTCCCCAGGGCAAG	1979
QY	1771	TCAGGCTCTCTCCCTCATAGCTCTTGTGTGCTATGTGCTTATAGGGTGGCTCTCATCTCCA	1833
Db	1980	TCAGGCTCTCTCCCTCATAGCTCTTGTGTGCTATGTGCTTATAGGGTGGCTCTCATCTCCA	2039
QY	1831	CCACTCTGGG-CCCTTGGGGAGAGACTGGGGAGGGGCGCTGGAGAGACCCCTGACGCTGG	1888
Db	2040	CCACTCTGGGCGCTTGGGGAGAGACTGGGGAGGGGCGCTGGAGAGACCCCTGACGCTGG	2099
QY	1890	AACCTGTATACAACAATTAAGGACAGTCTCACAGA	1923
Db	2100	AACCTGTATACAACAATTAAGGACAGTCTCACAGA	2133

AC ADCC31919,
XX
DT 18-DEC-2003 (first entry)
DE Human novel cDNA contig sequence, SEQ ID NO:2001.
XX
KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianemic; anticoagulant; thrombolytic; vulnery;
KW antitumor; osteoprotic; immunosuppressive; antiinflammatory; cytostatic;
KW gene therapy; chromosome 1p36.11-36.2; ss.
XX
OS Homo sapiens.
XX
PN M02003029271-A2.
XX
PD 10-APR-2003.
XX
PF 24-SEP-2002; 2002WC-US030474.
XX
FR 24-SEP-2001; 2001US-0324631P.
XX
FA (HYSEQ-) HYSEQ INC.
PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T,
PI Zhou P, Ghosh M, Wang D, Ma Y, Aundi V, Wang Z, Weng G,
PI Haley-Vicente D, Drmanac RT;
XX
DR WPI, 2003-371981/35.
XX
PT P-PSDB; ADCC32686.
XX
PT New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.
XX
PS Example 2; SEQ ID NO 2001; 11855p; English.
XX
XX The invention relates to 971 novel human cDNA sequences (ADCC29919-
CC ADCC30889)and the polypeptides they encode (ADCC0890-ADCC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of preventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (ADCC1861-ADCC2627) and the polypeptides encoded by the contigs (ADCC2622
CC ADCC3394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a human contig
CC sequence used in an example of the invention. Note: The sequence data for
CC this patent did not form part of the invention. Note: The sequence data for
CC obtained in electronic format directly from WIGO at
CC ftp.wigo.int/pub/published_pct_sequences.

XX Sequence 2142 BP; 451 A; 700 C; 568 G; 423 T; 0 U; 0 Other:
 SQ Query Match 84.0%; Score 1670; DB 10; Length 2142;
 Best Local Similarity 95.8%; Pred. No. 0;
 Matches 1815; Conservative 0; Mismatches 57; Indels 22; Gaps 9;

33 GCCCATCCCAACCAACCCCGGAGGCAACATCAAGACCTTGAAGAGCCCTATGA 92
 Db 259 GGCCTTCCCAACCCCGGAGGCAACATCAAGACCTTGAAGAGCCCTATGA 318
 Qy 93 GTTTCGGGTGAGCTGAGAGACTTCTCACTGAAGACATCATTTGTGACCACTCCAAACA 152
 Db 319 GTTTCGGGTGAGCTGAGAGACTTCTCACTGAAGACATCATTTGTGACCACTCCCA---A 375
 Qy 153 AACCAATCGAGGTGGGGGCTGAGAAAGTGGGGCTGAAGGAGCACTGTATCATGAAA 212
 Db 376 CAACCAATCGAGGTGGGG--CTGAGAGCTGGGGCTGAAGGAGCACTGTATCATGAAA 433
 Qy 213 CACCTTACGACTACAAAGTGCAGACTGGGAGGAGCTGAGACCTCCGAGCTGCGGTGA 272
 Db 434 TTGGCT-----CACAAAGTGCAG--CTGGGAGGAGCTGAGACCT---GAGGTGGTGA 482
 Qy 273 CCTCGCTACTGTGCGGAGAGAGAGACCTCTCACTAGTCCGGGACAGCTTCACTCCGATTA 332
 Db 483 CCTGG---CTTGGGAGAGAGAGAGCTCTCACTATCCGGGACAGCTTCACTCCGATTA 539
 Qy 333 CAGAACAGTTCAGAGAGCTTCCGAGAGGAGATCAAAAATCTGAGTCCCTCCCTCC 392
 Db 540 CAGAACAGTTCAGAGAGCTTCCGAGAGGAGATCAAAAATCTGAGTCCCTCCCTCC 599
 Qy 393 TTTTCCTGTCCCCCGCCCGAGCTGCGAGCAAGAGCTTGGCTAACCCCATTAACAAGC 452
 Db 600 TTTTCCTGTCCCCCGCCCGAGCTGCGAGCAAGAGCTTGGCTAACCCCATTAACAAGC 659
 Qy 453 TCCAGAGCATCTCAGCCAGAGTCTAGCCCGCCAGACCCCGAGCTGAGACATC 512
 Db 660 TCCAGAGCATCTCAGCCAGAGTCTAGCCCGCCAGACCCCGAGCTGAGACATC 719
 Qy 513 CTCCTCAACTAGGAGCTTCCCTCACTTATCAAGGAGAGGAGAGCTCCCTGGCTGAC 572
 Db 720 CTCCTCAACTAGGAGCTTCCCTCACTTATCAAGGAGAGGAGAGCTCCCTGGCTGAC 779
 Qy 573 ATGATGCCAGATTTGAGATTGGCTCCGCTCACTTATCCAGAGTACAGGGGCTGGGT 632
 Db 780 ATGATGCCAGATTTGAGATTGGCTCCGCTCACTTATCCAGAGTACAGGGGCTGGGT 839
 Qy 633 CAGGAGAGAGATCTTAAAGAACCACTGTGGGTCAAGGAGATGGGACAGAGACATA 692
 Db 840 CAGGAGAGAGATCTTAAAGAACCACTGTGGGTCAAGGAGATGGGACAGAGACATA 899
 Qy 693 TGGGCAAGCTTTCAGAGAGACAGACAGACAAACCTCTGTATCTATGAAGTCTTCAG 752
 Db 900 TGGGCAAGCTTTCAGAGAGACAGACAGACAAACCTCTGTATCTATGAAGTCTTCAG 959
 Qy 753 GGCAGAGGAGACAGAGAGCTTGAACCTCTTGGGCAAGGAGAGTGGAGAGACAGAGGA 812
 Db 960 GGCAGAGGAGACAGAGAGCTTGAACCTCTTGGGCAAGGAGAGTGGAGAGAGAGAGGA 1019
 Qy 813 AGGTTCAGGCAAGGAGTCTTATCTTAAAGTGAATCTATTTGCCGAGGGCTCGCAAGGCC 872
 Db 1020 AGGTTCAGGCAAGGAGTCTTATCTTAAAGTGAATCTATTTGCCGAGGGCTCGCAAGGCC 1079
 Qy 873 AAGAGAGACAGCGCTGACGATTAACCTCCCTCTAACCGCTTCAAGGCCCAAGCCAGC 932
 Db 1080 AAGAGAGACAGCGCTGACGATTAACCTCCCTCTAACCGCTTCAAGGCCCAAGCCAGC 1139
 Qy 933 GAGCAGGCTGCTGCGCAACCCCGTGGCCCAAGCTGTGCTGCAAGGAGAGCCAT 992
 Db 1140 GAGCAGGCTGCTGCGCAACCCCGTGGCCCAAGCTGTGCTGCAAGGAGAGCCAT 1199
 Qy 993 GGCACATCTGTATATAGATGGGGTTTTTCCAAATACAGTGGTCTGTGAAAACTGCATGA 1052

Db 1200 GGCACATCTGTATATAGATGGGGTTTTTCCAAATACAGTGGTCTGTGAAAACTGCATGA 1259
 Qy 1053 AACTCTGCGCTCTGCGGCTGCTGGGAGCTTCAGGCAAGGCAAGTGGGGTGGGGGTG 1112
 Db 1260 AACTCTGCGCTCTGCGGCTGCTGGGAGCTTCAGGCAAGGCAAGTGGGGTGGGGGTG 1319
 Qy 1113 GGGCTGCTCTTCCCTCCCAAGGAGCTGTGTTCTGAGGGCTGCTCCATGAGAGAGG 1172
 Db 1320 GGGCTGCTCTTCCCTCCCAAGGAGCTGTGTTCTGAGGGCTGCTCCATGAGAGAGG 1379
 Qy 1173 ATCACTTAACAGAGATGGAAGCCAGGCAATGATGGGGCTTTGGGTCTCGAGGTTGAC 1232
 Db 1380 ATCACTTAACAGAGATGGAAGCCAGGCAATGATGGGGCTTTGGGTCTCGAGGTTGAC 1439
 Qy 1233 CCGAGCTTCTTCCCACTCCCTCC--GGAGTGAAGTCTCCATCCATCCCTCCCTTAA 1291
 Db 1440 CCGAGCTTCTTCCCACTCCCTCCGGAGAGTGAAGTCTCCATCCATCCCTCCCTTAA 1499
 Qy 1292 TCTATGAATCTAATAGGCTGGTGTGTATAA--ACACACCTTATCGTTGCTTCAAT 1350
 Db 1500 TCTATGAATCTAATAGGCTGGTGTGTATAACACACACACCTTATCGTTGCTTCAAT 1559
 Qy 1351 ACTAGCATTTACATTTGGTGTGAGCCCAATTCAGAGCTTTCTCAATCAATTTTACATC 1410
 Db 1560 ACTAGCATTTACATTTGGTGTGAGCCCAATTCAGAGCTTTCTCAATCAATTTTACATC 1619
 Qy 1411 TCCATTTTCAATTAAGGGGAAACATCCCGAGCACTGAAGTGTGCTTGTCACTGA 1470
 Db 1620 TCCATTTTCAATTAAGGGGAAACATCCCGAGCACTGAAGTGTGCTTGTCACTGA 1679
 Qy 1471 GGTTAGATCTGAACCCAGGGTGTCAAGCTGCTTCAACTCCCACTGTGGGCACTGA 1530
 Db 1680 GGTTAGATCTGAACCCAGGGTGTCAAGCTGCTTCAACTCCCACTGTGGGCACTGA 1739
 Qy 1531 GGAATATTTCCCTCATTTCTACCTCTTAAGGCTATGCAACCTCCCAAGCTTTCAGC 1590
 Db 1740 GGAATATTTCCCTCATTTCTACCTCTTAAGGCTATGCAACCTCCCAAGCTTTCAGC 1799
 Qy 1591 TGGGAGATGGGGAGTGCATAGAAAAGCCCATCTCCATGTGGATAGGAGCTTCC 1650
 Db 1800 TGGGAGATGGGGAGTGCATAGAAAAGCCCATCTCCATGTGGATAGGAGCTTCC 1859
 Qy 1651 ATCAGCCTTAACCTTGGGAAATGCTGCTGCCCCCAGTGACTGTTGGTTCGCTCCAC 1710
 Db 1860 ATCAGCCTTAACCTTGGGAAATGCTGCTGCCCCCAGTGACTGTTGGTTCGCTCCAC 1919
 Qy 1711 ATACAGAGAGGAGTGAAGGAGAAAGGTGGGTCTCAGTTAGCAGGGGTCCCAAGGCAAG 1770
 Db 1920 ATACAGAGAGGAGTGAAGGAGAAAGGTGGGTCTCAGTTAGCAGGGGTCCCAAGGCAAG 1979
 Qy 1771 TCAGCTCTCTCCCTCCATGCTCTCTTGTGATGTGTGCTTAAAGGTGGCTCTCACTCCCA 1830
 Db 1980 TCAGCTCTCTCCCTCCATGCTCTCTTGTGATGTGTGCTTAAAGGTGGCTCTCACTCCCA 2039
 Qy 1831 CCACTCTGGG--CCCTTGGGGAGAGACTGGGAGAGGGGCGTGGGAGAGCCCTGA 1889
 Db 2040 CCACTCTGGGCGCTTGGGGAGAGACTGGGAGAGGGGCGTGGGAGAGCCCTGA 2039
 Qy 1890 AACCTGTATACAAATTAAGAGAGCTCTACAGA 1923
 Db 2100 AACCTGTATACAAATTAAGAGAGCTCTACAGA 2133

RESULT 10
 ADM19231
 ID ADM19231 standard; cDNA; 2220 BP.
 XX
 XX ADM19231;
 XX AC
 XX DT 20-MAY-2004 (first entry)
 XX DE
 XX Novel human channel/transporter gene #28.

KM de; gene; immunosuppressive; antiarthritic; anti-rheumatic;
KM antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KM neurotropic; neuroprotective; antibacterial; virucide; fungicide;
KM ophthalmological; gene therapy; channel/transporter protein;
KM rheumatoid arthritis; neoplasm; cardiac arrest; cerebrovascular disorder;
KM cerebral ischemia; angiogenesis; nervous system disorder;
KM Alzheimer's disease; ocular disorder; corneal infection; wound healing;
KM epithelial cell proliferation; skin aging; sunburn; transplantation;
KM chemotaxis; food additive.
XX
OS Homo sapiens.
XX
EN W0200154472-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001307.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205555P.
PR 07-JUN-2000; 2000US-0209457P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217486P.
PR 14-JUL-2000; 2000US-0218230P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228928P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229345P.
PR 01-SEP-2000; 2000US-0229346P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.

PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 02-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 05-DEC-2000; 2000US-0251990P.

PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM,
XX WPI, 2001-476159/51.
DR P-PSDE; ADM19710.
XX
PT Isolated nucleic acid molecule encoding a channel/transporter protein is
XX used in preventing, treating or ameliorating a medical condition.
XX
PS Claim 1; SEQ ID NO 38; 809pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule encoding a
XX channel/transporter protein or sequences at least 95% identical to a
XX these. The nucleic acids and proteins encoded by them are used to
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
XX in diagnosing a pathological condition or susceptibility to a
XX petriological condition. The antibodies to the proteins can also be used
XX in alleviating symptoms associated with the disorders and in diagnostic
XX immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
XX (ELISA). Disorders which are diagnosed or treated include autoimmune
XX diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
XX neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
XX arrest, cerebrovascular disorders e.g. cerebral ischemia, angiodenesis,
XX nervous system disorders e.g. Alzheimer's disease, infections caused by
XX bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
XX The polypeptides can also be used to aid wound healing and epithelial
XX cell proliferation, to prevent skin aging due to sunburn, to maintain
XX organs before transplantation, for supporting cell culture of primary
XX tissues, to regenerate tissues and in chemotaxis. The polypeptides can
XX also be used as a food additive or preservative to increase or decrease
XX storage capabilities. This sequence corresponds to a gene of the
XX invention.
XX
SQ Sequence 2220 BP; 461 A; 718 C; 597 G; 429 T; 0 U; 15 Other;
Query Match 83.9%; Score 1668; DB 5; Length 2220;
Best Local Similarity 95.6%; Pred. No. 0;
Matches 1810; Conservative 5; Mismatches 57; Indels 22; Gaps 9;
QY 33 GCCCATCCCAACCAAGCCCGGTGGGAGCAATCAAGACCTTAGAGAGCGCTATGA 92
Db 313 GGGCTTCCCAAGCCCGCGGTGGGAGCAATCAAGACCTTAGAGAGCGCTATGA 372
QY 93 GTTTGGGTGGAGCGTGAAGATCTTCACTGGAAGACATCATTTGACCCCTCAACA 152
Db 373 GTTTGGGTGGAGCGTGAAGATCTTCACTGGAAGACATCATTTGACCCCTCAACA 429
QY 153 AACCAATCGAGTGGCGGGCTGAGAAAGCTGGCGGCTGAACGACACTGATGAAA 212
Db 430 CAACCAATCGAGTGGCGGGCTGAGAAAGCTGGCGGCTGAACGACACTGATGAAA 487
QY 213 CACCTTAGACCTACCAAGTGGAGACTGGCGGAGAGCTGAGACCCCGAGCGTGTGA 272
Db 488 TTGCGT-----CAAGATGCCAG-CTGCCGAGAGAGCTGAGACCC--GACGTCGGTGA 536
QY 273 CTTGGGCTACTGTGGGAGAGAGCGAGCAAGCTCACTAGTCCGGGAGCGGTGATAC 332
Db 537 CTTGGG---CTTGGGAGAGAGCGAGCAAGCTCACTAGTCCGGGAGCGGTGATAC 593
QY 333 CAGAACAGCTCCAGAGACTTTCGCGAGCGAGATCAAAATCTGAGTCTCTCCCTTCC 392
Db 594 CAGAACAGCTCCAGAGACTTTCGCGAGCGAGATCAAAATCTGAGTCTCTCCCTTCC 653

QY 393 TTTCCTGTGCCCCCGCCCGCCAGAGCAAGAGCTGCTTAACCCCTTACACAGC 452
Db 654 TTTCCTGTGCCCCCGCCCGCCAGAGCAAGAGCTGCTTAACCCCTTACACAGC 713
QY 453 TCGAGCATCTGAGCCAGGTTCTAGCCCGCCAGAGCAAGAGCTGAGTGAACCATC 512
Db 714 TCGAGCATCTGAGCCAGGTTCTAGCCCGCCAGAGCAAGAGCTGAGTGAACCATC 773
QY 513 CTCCCAACTAGGGGCTTCCACTATTCAGAGGAGAGGAGAGTCTCTGGCTGACAC 572
Db 774 CTCCCAACTAGGGGCTTCCACTATTCAGAGGAGAGGAGAGTCTCTGGCTGACAC 833
QY 573 ATGATGCCAGATTTCAGATTGAGCTCCCTCACTTATCAGAGTACAGAGGCTGGAGT 632
Db 834 ATGATGCCAGATTTCAGATTGAGCTCCCTCACTTATCAGAGTACAGAGGCTGGAGT 893
QY 633 CAGGGAAGAGATCTTAAGAACCTCAGTGGGTCAGGGGAATGGGACACAGACATTA 692
Db 894 CAGGGAAGAGATCTTAAGAACCTCAGTGGGTCAGGGGAATGGGACACAGACATTA 953
QY 693 TGGCAAGCTCTGAGAGACAGACAGACAGACAAACCTGATCTATGAAGTCTGACAG 752
Db 954 TGGCAAGCTCTGAGAGACAGACAGACAGACAAACCTGATCTATGAAGTCTGACAG 1013
QY 753 GGCAGAGGGAGCAGAGGAGCTTGAACCTCTTGGCCAAAGGGAGTGGAGAGACAGAG 812
Db 1014 GGCAGAGGGAGCAGAGGAGCTTGAACCTCTTGGCCAAAGGGAGTGGAGAGACAGAG 1073
QY 813 AGGTCAAGGCAAGGGGCTATCTATGAGAGAACTAATGGGAGGCTCAGCAAGGCG 872
Db 1074 AGGTCAAGGCAAGGGGCTATCTATGAGAGAACTAATGGGAGGCTCAGCAAGGCG 1133
QY 873 AAGAGAGACAGGCGTGAAGCTTCCCTCACTCAAGCTCCAGGCGCCAGCGCAGC 932
Db 1134 AAGAGAGAGAGCGTGAAGCTTCCCTCACTCAAGCTCCAGGCGCCAGCGCAGC 1193
QY 933 GAGAGAGGCTGCGCCAGCCCGCTGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAG 992
Db 1194 GAGAGAGGCTGCGCCAGCCCGCTGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAG 1253
QY 993 GCCACATCTGTATATAGATGGGTTTTCATATACAGCTGTTGCGAATAAATCGATGA 1052
Db 1254 GCCACATCTGTATATAGATGGGTTTTCATATACAGCTGTTGCGAATAAATCGATGA 1313
QY 1053 AACTCTGCGGCTGCGCCAGGCTGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAG 1112
Db 1314 AACTCTGCGGCTGCGCCAGGCTGCGCCAGGAGGAGGAGGAGGAGGAGGAGGAGG 1373
QY 1113 GGGCTGCTCTTCTCCCTCCAGAGGCTGTGTGTTGAGGGCTGCTCCATGAGAGCAG 1172
Db 1374 GGGCTGCTCTTCTCCCTCCAGAGGCTGTGTGTTGAGGGCTGCTCCATGAGAGCAG 1433
QY 1173 ATCACTTAACAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1232
Db 1434 ATCACTTAACAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1493
QY 1233 CCGAGCTCTGCGACCTTCCCTCC-GGCACTCAGCTCTGATCATCTCCCTCTTAA 1291
Db 1494 CCGAGCTCTGCGACCTTCCCTCC-GGCACTCAGCTCTGATCATCTCCCTCTTAA 1553
QY 1292 TCTATGAATCTATAGGCTGAGTGTGTGAACA-AACAACCTCATCTGTGTCTTCAAT 1350
Db 1554 TCTATGAATCTATAGGCTGAGTGTGTGTGAACAACAACCTCATCTGTGTCTTCAAT 1613
QY 1351 ACTCAGCATTTACATTTGTTGAGGCGCAATTCAGAGCTTTCTCAATCAGATTTCAATC 1410
Db 1614 ACTCAGCATTTACATTTGTTGAGGCGCAATTCAGAGCTTTCTCAATCAGATTTCAATC 1673
QY 1411 TCCATTTTCAATTAACGGGGAACATCCCGAGAGCCTAGTGTGCTGCTTTGTCACTGAA 1470
Db 1674 TCCATTTTCAATTAACGGGGAACATCCCGAGAGCCTAGTGTGCTGCTTTGTCACTGAA 1733

Qy	1471	GGTTAGATCTCAACCCAGGAGGTCAACNGCTGCTCTCAATCCCACTCTGGGCACTGA	1530
Db	1734	GGTTAGATCTCAACCCAGGAGGTCTCAACMGCTGCTCTCAATCCCACTCTGGGCACTGA	1793
Qy	1531	GGAGTATTTCCCTCATTTCTAACCCTCTTAAGGCTATAGCAACCCCTTCCACGCTTCCAGC	1590
Db	1794	GGAGTATTTCCCTCATTTCTAACCCTCTTAAGGCTATAGCAACCCCTTCCACGCTTCCAGC	1853
Qy	1531	TGGGGGATGGGGGAGGTATAGAAAAGCCCCCATCTCCCATCTGGGATAGGSAACCTTCC	1650
Db	1854	TGGGGGATGGGGGAGGTATAGAAAAGCCCCCATCTCCCATCTGGGATAGGSAACCTTCC	1913
Qy	1651	ATCAGCCCTTAAACCCCTGGGAAATGCTGCTGCCCCCAGTACTCTTGATTTGCTCCAC	1710
Db	1914	ATCAGCCCTTAAACCCCTGGGAAATGCTGCTGCCCCCAGTACTCTTGATTTGCTCCAC	1973
Qy	1711	ATPACGAAGCAAGGCTGAGGGGGAAGGGTGTCTTAATTAGCAGGGGCTCCCAAGGGAAG	1770
Db	1974	ATPACGAAGCAAGGCTGAGGGGGAAGGGTGTCTTAATTAGCAGGGGCTCCCAAGGGAAG	2033
Qy	1771	TCAGGCTCTCCCTCCATGCTCTCTCTGCTGCACTGTGCTTAAAGGTGAGCTCTCACTCCA	1830
Db	2034	TCAGGCTCTCTCCCTCCATGCTCTCTCTGCTGCACTGTGCTTAAAGGTGAGCTCTCACTCCA	2093
Qy	1831	CCACTCTTGGG-CCCTTGGGGGAGAGACTGGGGAGGGGGCCGTGGAGAGACCTTGAAGCTGG	1888
Db	2094	CCACTCTTGGG-CCCTTGGGGGAGAGACTGGGGAGGGGGCCGTGGAGAGACCTTGAAGCTGG	2153
Qy	1890	AACTGTATATACAAATAAAGGCAACTCTCACAGA	1923
Db	2154	AACTGTATATACAAATAAAGGCAACTCTCACAGA	2187

RESULT 11

ID ADQ22267 standard; DNA; 3503 BP.

AC ADQ22267;

DT 26-AUG-2004 (first entry)

DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 5087.

XX soft tissue sarcoma; cyrostatic; gene therapy; vaccine; screening; human;

KW ds.

XX Homo sapiens.

OS

XX WO2004048938-A2.

PN

XX 10-JUN-2004.

PD

XX 26-NOV-2003; 2003WO-US038193.

PE

XX 26-NOV-2002; 2002US-0429739P.

PR

XX (PROT-) PROTEIN DESIGN LABS INC.

PA

XX Aziz N, Ginsburg WM, Zlotnik A;

PI

XX MPI; 2004-441208/41.

DR

XX

PT Early detection of soft tissue sarcoma comprises determining expression

PT of a gene in a first soft tissue sample and a normal soft tissue sample

PT and comparing the gene expression, also useful in treating soft tissue

PT sarcoma.

PS

XX

PS Example 2; SEQ ID NO 5087; 210pp; English.

XX

CC The invention relates to a novel method for detecting soft tissue sarcoma

CC which comprises obtaining a first soft tissue sample from an individual

CC and a normal soft tissue sample from the same or different individual,

CC determining the expression of a gene in both samples and comparing the

Best Local Similarity 95.6%; Pred. No. 0;
Matches 1808; Conservative 0; Mismatches 61; Indels 22; Gaps 9;

```
QY 33 GCCCATCCCAACGAGCCCGGTGGGAGGCAACATCAAGACCTTAGAGAGCGCTATGA 92
Db 539 GGCCTTCCAGCCCGCCCGGTGGGAGGCAACATCAAGACCTTAGAGAGCGCTATGA 598
QY 93 GTTGGGTGAGAGTGAAGATCTTCACTGAAGACATCATGTATCAACACCTCCAAAGA 152
Db 599 GTTGGGTGAGAGTGAAGATCTTCACTGAAGACATCATGTATCAACACCTCCAAAGA 655
QY 153 AACCAATCATGAGGTGGGAGGAGTGAAGAAAGCTGGCGTGAACGGAACATGTACATGAA 212
Db 656 CAACCAATCATGAGGTGGGAGGAGTGAAGAAAGCTGGCGTGAACGGAACCGTCAATGAACAC 713
QY 213 CACTTACGACTACACAAAGTCCAGACTGCGGAGAGAGCTGAGACCCCGAGCGTGGGTGA 272
Db 714 TTGGCT-----CACAAAGTCCAG-CTGCCGAGAGAGTGAACCC--GACGTGGTGA 762
QY 273 CCTGGCTACTGTGGGAGAGAGGAGAGCTCACTAGTCCGGGAGAGCGCTCAACCCGATGA 332
Db 763 CTTGG---CTGTGGGAGAGAGGAGAGCTCACTAGTCCGGGAGAGCGCGCTCAACCCGATGA 819
QY 333 CAGAACAGTCCAGAGAGCTTTCCGAGCGAGATCAAAATCTGATGCTCTTCCCTTCCC 392
Db 820 CAGAACAGTCCAGAGAGCTTTCCGAGCGAGATCAAAATCTGATGCTCTTCCCTTCCC 879
QY 393 TTTTCCCTGTCCCGCCCGCCAGCGCTGCGAGCAAGCGTGTAAACCCCACTTACACAGC 452
Db 880 TTTTCCCTGTCCCGCCCGCCAGCGCTGCGAGCAAGCGTGTAAACCCCACTTACACAGC 939
QY 453 TCCAGAGCATCTCAGCCAGGTTCTAGCCCCCAAGCAAGCCCAAGACCCAGAGTGAACATC 512
Db 940 TCCAGAGCATCTCAGCCAGGTTCTAGCCCCCAAGCAAGCCCAAGAGTGAACATC 999
QY 513 CTCCCAACTAGAGGCGCTTCACTATTCAGAGGAGAGGAGAGTCCCTGGGCTGACAC 572
Db 1000 CTCCCAACTAGAGGCGCTTCACTATTCAGAGGAGAGGAGAGTCCCTGGGCTGACAC 1059
QY 573 ATGATGCCAGATTTAGATTTGGCTCCGTCACTTATCCAGAGTCAAGGGCTGGAGT 632
Db 1060 ATGATGCCAGATTTAGATTTGGCTCCGTCACTTATCCAGAGTCAAGGGCTGGAGT 1119
QY 633 CAGGAGAGAAAGATCTTAAAGAACCCACTGTGGGTCAAGGGAGATGGGACCAAGACATA 652
Db 1120 CAGGAGAGAAAGATCTTAAAGAACCCACTGTGGGTCAAGGGAGATGGGACCAAGACATA 1179
QY 693 TGGGCAAGCTCTGAGGAGAGAGAGAGCAAGAACCCCTGATGTATAGAGTCTTGGCAG 752
Db 1180 TGGGCAAGCTCTGAGGAGAGAGAGAGAGCAAGAACCCCTGATGTATAGAGTCTTGGCAG 1239
QY 753 GGCAGAGGAGACCAAGGAGCTTGAACCCCTTGTGGCAAGGGAGAGTGGAGAGACAGAGGA 812
Db 1240 GGCAGAGGAGACCAAGGAGCTTGAACCCCTTGTGGCAAGGGAGAGTGGAGAGACAGAGGA 1299
QY 813 AGGTCAAGAGGAGAGAGTGGCTTATAGTGAATTAATTGCCGAGGGGCTCAGAGAGGC 872
Db 1300 AGGTCAAGAGGAGAGAGTGGCTTATAGTGAATTAATTGCCGAGGGGCTCAGAGAGGC 1359
QY 873 AAGAGAGAGACAGCGGTGACGTTAACTTCCCTCTACAGAGCTTCAAGCCCAAGCCAGC 932
Db 1360 AAGAGAGAGACAGCGGTGAGTTAACTTCCCTCTACAGAGCTTCAAGCCCAAGCCAGC 1419
QY 933 GAGCAGGTGCTGCTCCCAACCCCGTGGCCCAAGCAGAGTGGTGTGCCAGAGGAGAGCAT 992
Db 1420 GAGCAGGTGCTGCTCCCAACCCCGTGGCCCAAGCAGAGTGGTGTGCCAGAGGAGAGCAT 1479
QY 993 GCCACATCTGTATATAGATGGGGTTTTTCCAAATCAGAGTGGTGTGTAATAAAGTGCATGA 1052
Db 1480 GCCACATCTGTATATAGATGGGGTTTTTCCAAATCAGAGTGGTGTGTAATAAAGTGCATGA 1539
QY 1053 AACTCTGCGGTCTGCGGCTGTGGGGCTTCCAGGCAAGGCCAAGTGGGGTTGGGGGTG 1112
```

```
Db 1540 AACTCTGCGGTCTGCGGCTGTGGGGCTTCCAGGCAAGGCCAAGTGGGGTTGGGGGTG 1599
QY 1113 GGGCTGTGCTTCTTCTCCCTCCCAAGGCGCTGTGTTCTNAGGGGCTGTCCCAAGCAGACAG 1112
Db 1600 GGGCTGTGCTTCTTCTCCCTCCCAAGGCGCTGTGTTCTNAGGGGCTGTCCCAAGCAGACAG 1659
QY 1173 ATCACTTAACAGAGATGGAAGCCAGGAGATGAGATGGGGCTTTGGGTCTCTCAGGTGAGC 1232
Db 1660 ATCACTTAACAGAGATGGAAGCCAGGAGATGAGATGGGGCTTTGGGTCTCTCAGGTGAGC 1719
QY 1233 CCGAGCTTGTGCCACTTCCCTTCC--GGAGTCAAGCTCTGCATTCATCCCTCTTTAA 1291
Db 1720 CCGAGCTTGTGCCACTTCCCTTCCCGGAGAGTCAAGCTCTGCATTCATCCCTCTTTAA 1779
QY 1292 TCTATGAATCTATAGGCTCGGTGTGTATACA-ACACACCCCTATCGTTGTCTTCAAT 1350
Db 1780 TCTATGAATCTATAGGCTCGGTGTGTATACAACACACCCCTATCGTTGTCTTCAAT 1839
QY 1351 ACTCAGCATTTACATTTGATGGCCAAATTCAGAGCTTTCTCAATCAATCAATTCAATC 1410
Db 1840 ACTCAGCATTTACATTTGATGGCCAAATTCAGAGCTTTCTCAATCAATCAATTCAATC 1899
QY 1411 TCCATTTTCAATTAAGGGGAAACATCCCGAGGCACTGAGTGTGTTGTGATCACTGAA 1470
Db 1900 TCCATTTTCAATTAAGGGGAAACATCCCGAGGCACTGAGTGTGTTGTGATCACTGAA 1959
QY 1471 GGTATGATCTGAACCCAGGGTGTCAACAGCTGTCTCACTCCCACTCTGGGACATGA 1530
Db 1960 GGTATGATCTGAACCCAGGGTGTCAACAGCTGTCTCACTCCCACTCTGGGACATGA 2019
QY 1531 GGAATATTTCCCTCATTTACTCTCTTAAGGCTATGACCCCTCCCAAGCTTCCAGC 1590
Db 2020 GGAATATTTCCCTCATTTACTCTCTTAAGGCTATGACCCCTCCCAAGCTTCCAGC 2079
QY 1591 TGGGGAGATGGGGAGTCAATAGAAAGGCCCATCTCCATCTGGAGATAGGAGACTTCC 1650
Db 2080 TGGGGAGATGGGGAGTCAATAGAAAGGCCCATCTCCATCTGGAGATAGGAGACTTCC 2139
QY 1651 ATCAGCTTTAACCTTGGGAAATGCTGTCTGCTGCCAGTGAAGCTTTGGTTGCTTCCAC 1710
Db 2140 ATCAGCTTTAACCTTGGGAAATGCTGTCTGCTGCCAGTGAAGCTTTGGTTGCTTCCAC 2199
QY 1711 ATACAGAGCAAGGATGGAGGAGAGGAGTGGGCTCAGTTAGAGAGGAGTCCCAAGGCAAG 1770
Db 2200 ATACAGAGCAAGGATGGAGGAGAGGAGTGGGCTCAGTTAGAGAGGAGTCCCAAGGCAAG 2259
QY 1771 TCAGCTCTCTCCCTCATCTCTGTGTCAAGTGTAGGCTTGGAGTGGCTCTCACTCCCA 1830
Db 2260 TCAGCTCTCTCCCTCATCTCTGTGTCAAGTGTAGGCTTGGAGTGGCTCTCACTCCCA 2319
QY 1831 CCAGTCTGGG--CCCTTGGGGAGAGCTGGGAGGGGGCCGTGGGAGAGCCTGAAGCTGG 1889
Db 2320 CCAGTCTGGG--CCCTTGGGGAGAGCTGGGAGGGGGCCGTGGGAGAGCCTGAAGCTGG 2379
QY 1890 AACCTGTATACAAATTAAGAGCAAGTCTCAC 1920
Db 2380 AACCTGTATACAAATTAAGAGCAAGTCTCAC 2410
```

RESULT 13
AAK94832
ID AAK94832 standard; cDNA; 2701 BP.
XX
XX AAK94832;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human full-length cDNA, SEQ ID NO: 3983.
XX
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
XX Homo sapiens.
XX

PN EP1130094-A2.
XX
XX 05-SEP-2001.
XX
XX 07-JUL-2000; 2000EP-00114089.
XX
XX 08-JUL-1999; 99JP-00194486.
XX 11-JAN-2000; 2000JP-00118774.
XX 02-MAY-2000; 2000JP-00183765.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX WPI, 2001-524255/58.
XX
XX P-P-SDB; AAM93873.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX
XX
XX Claim 8; SEQ ID NO 3983; 1380bp + Sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5' and 3' ends of the cDNA molecules have
CC been determined. Primers for synthesizing the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesized by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a full length human cDNA of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in CD-ROM format directly
CC from EPO
XX
SQ Sequence 2701 BP, 557 A, 867 C, 771 G, 506 T, 0 U, 0 Other;
Query Match 83.5%; Score 1658.8; DB 4; Length 2701;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1806; Conservative 0; Mismatches 59; Indels 23; Gaps 9;
QY 39 CCCAACCAAGCCCCGCTGGGAGAGCAATCAAGACCTTAGAGAGCGCTTAGATTGGC 98
DB 833 CAAGAGCCGCGCGGGGAGGAGCAATCAAGACCTTAGAGAGCGCTTAGATTGGC 892
QY 99 GGTGAGAGTGAAGACTTCTCACTGAAAGACATCTTGTCAACCACTCCAAACAAACCA 158
DB 893 GGTGAGAGTGAAGACTTCTCACTGAAAGACATCTTGTCAACCACTCCAAACAAACCA 943
QY 159 CATGAGAGTGAAGACTTCTCACTGAAAGACATCTTGTCAACCACTCCAAACAAACCA 218
DB 950 CATGAGAGTGAAGACTTCTCACTGAAAGACATCTTGTCAACCACTCCAAACAAACCA 1007
QY 219 AGGACTACAGAGTGAAGACTTCTCACTGAAAGACATCTTGTCAACCACTCCAAACAAACCA 278
DB 1008 -----CAAGAGTGAAGACTTCTCACTGAAAGACATCTTGTCAACCACTCCAAACAAACCA 1056
QY 279 CTACTGTGAG 338
DB 1057 ---CTGTGAG 1113
QY 339 AGGTGAG 398
DB 1114 AGGTGAG 1173
QY 399 TGTGAG 458
DB 1174 TGTGAG 1233
QY 459 ACATCTCAG 518
DB 1234 ACATCTCAG 1293

QY 519 AACTAGAGCCCTCCACTCTATTCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 578
DB 1294 AACTAGAGCCCTCCACTCTATTCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1353
QY 579 CCGAGATTGAGATTGGGCTCCGCTCACTTAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638
DB 1354 CCGAGATTGAGATTGGGCTCCGCTCACTTAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1413
QY 639 AGGAGATCTTAAG 698
DB 1414 AGGAGATCTTAAG 1473
QY 699 AGCTTCGAG 758
DB 1474 AGCTTCGAG 1533
QY 759 GGGACAG 818
DB 1534 GGGACAG 1593
QY 819 CAGGCAAG 878
DB 1594 CAGGCAAG 1653
QY 879 AG 938
DB 1654 AG 1713
QY 939 GCT 998
DB 1714 GCT 1773
QY 999 TCTGATATGATGAG 1058
DB 1774 TCTGATATGATGAG 1833
QY 1059 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1118
DB 1834 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1893
QY 1119 GTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1178
DB 1894 GTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1953
QY 1179 TAAAGAGATGAG 1238
DB 1954 TAAAGAGATGAG 2013
QY 1239 TTCTTGGCACTTCCCTCC - GGGAGTCAAGCTTCCATCCATCCCTCTTTAATATATG 1297
DB 2014 TTCTTGGCACTTCCCTCC - GGGAGTCAAGCTTCCATCCATCCCTCTTTAATATATG 2073
QY 1298 AATCTAAG 1356
DB 2074 AATCTAAG 2133
QY 1357 CATTAACATGAG 1416
DB 2134 CATTAACATGAG 2193
QY 1417 TTCAATTAAG 1476
DB 2194 TTCAATTAAG 2253
QY 1477 ATCTGAAG 1536
DB 2254 ATCTGAAG 2313
QY 1537 TTTCCTCTCAATTTACTCTCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1596
DB 2314 TTTCCTCTCAATTTACTCTCTTAAG 2373

XX	Human; diagnostic; drug screening; forensics; gene mapping;
KW	biomimetic; biomimetic; Parkinson's disease; Alzheimer's disease;
KM	neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KM	ulcers; osteoporosis; autoimmune disease; cancer;
KM	molecular weight marker; food supplement; antiparkinsonian; noctropic;
KM	neuroprotective; antianemic; anticoagulant; thrombolytic; vulnery;
KM	antifungal; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW	gene therapy; chromosome 1p36.11-36.2; gene; ss.
XX	
OS	Homo sapiens.
PN	WO2003029271-A2.
XX	
PD	10-APR-2003.
XX	
PF	24-SEP-2002; 2002WO-US030474.
XX	
PR	24-SEP-2001; 2001US-0324631P.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
P1	Zhang TY, Zheng J, Ren F, Xue AJ, Zhao QA, Wang J, Mehrman T;
XX	Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
P1	Halley-Vicente D, Drmanac RT;
XX	
XX	WPI; 2003-371981/35.
XX	P-PSDB; ADC030955.
XX	
PT	New polynucleotide and polypeptide useful for diagnosing, preventing or
PT	treating conditions such as neurodegenerative diseases, anemias, platelet
P1	disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT	cancer.
XX	
PS	Claim 1; SEQ ID NO 66; 1185bp; English.
XX	
CC	The invention relates to 971 novel human cDNA sequences (ADC229919-
CC	ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
CC	invention also relates to nucleic acid sequences over 99% identical with
CC	the novel human cDNAs. The invention additionally encompasses expression
CC	vectors and host cells comprising a nucleic acid of the invention; the
CC	recombinant production of a polypeptide of the invention; an antibody
CC	against a polypeptide of the invention; a method of detecting
CC	polynucleotides or polypeptides of the invention; and methods of
CC	identifying a compound which binds to a polypeptide of the invention. The
CC	invention further discloses methods of preventing, treating or
CC	ameliorating a medical condition; kits comprising polynucleotide probes
CC	and/or monoclonal antibodies for carrying out the methods of the
CC	invention; methods for the identification of compounds that modulate the
CC	expression or activity of the polynucleotide and/or polypeptide; and 767
CC	coding sequences corresponding to the cDNA sequences of the invention
CC	(ADC31861-ADC32627) and the polypeptides encoded by the cDNAs (ADC32628
CC	-ADC33394). The nucleic acids and polypeptides of the invention are
CC	useful in diagnostics, drug screening, forensics, gene mapping, in the
CC	identification of mutations responsible for genetic disorders or other
CC	traits, for assessing biodiversity, and in producing many other types of
CC	data and products dependent on DNA and amino acid sequences. They are
CC	also used for treating diseases such as Parkinson's disease, Alzheimer's
CC	disease and other neurodegenerative diseases, anaemia, platelet
CC	disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC	cancer. The nucleic acids may also be used as hybridisation probes or
CC	primers, and in the recombinant production of a protein. The polypeptides
CC	are also useful in generating antibodies, as molecular weight markers,
CC	and as food supplements. The present sequence represents a specifically
CC	claimed human cDNA sequence of the invention. Note: The sequence data for
CC	this patent did not form part of the printed specification, but was
CC	obtained in electronic format directly from Wipo at
CC	ftp.wipo.int/pub/published_pct_sequences.
XX	
-SQ	Sequence 2783 BP; 586 A; 880 C; 781 G; 536 T; 0 U; 0 Other;

Query Match	82.9%	Score 1648.2;	DB 10;	Length 2783;
Best Local Similarity	95.3%	Pred. No. 0;		
Matches 1812/2	Conservative 0;	Mismatches 60;	Indels 29;	Gaps 10;

QY 33 GCCATCCCAACCAAGCCCGGTGGGAGGCAATCAAGACCTTAGAGACGGCTATGA 92
Db 894 GGCCCTTCCACCCCGCCCGGTGGGAGGCAATCAAGACCTTAGAGACGGCTATGA 953
QY 93 GTTGGCGGTGAGCGTGAAGAGCTTCTCACTGAGAGACATCACTTGTCAACCTCCACACA 152
Db 954 GTTGGCGGTGAGCGTGAAGAGCTTCTCACTGAGAGACATCACTTGTCAACCTCCACA 1010
QY 153 AACCAATCGAGGTGCGGAGCTGAGAAAGCTGAGCGGTGAAACGACACTGTACATGAAA 212
Db 1011 CAACCAATCGAGGTGCGGAGCTGAGAAAGCTGAGCGGTGAAACGACACTGTACATGAAA 1068
QY 213 CACCTTAGACTACACAAAGTCCAGACTGCGCGGAGAGAGTGTGAGACCCGAGCGGTGTA 272
Db 1069 TTGCGT-----CACAAGTGCAG-CTGCGGAGAGAGTGTGAGACCC--GACGTGGTGA 1117
QY 273 CCTGCGCTACTGTGCGGAGAGAGACAGCTCACTAGTCCGCGGAGCGGTGACCGGATA 332
Db 1118 CTTGCGGAGAGAGAGCGGAGCGGTCACTATCCGCGGAGCGGTGACCGGATA 1174
QY 333 CAGAACAGTTCAGAGAGCTTCCGAGCGAGATCAAAATGTAGTCCCTTCCCTTCC 392
Db 1175 CAGAACAGTTCAGAGAGCTTCCGAGCGAGATCAAAATGTAGTCCCTTCCCTTCC 1234
QY 393 TTTCCGCTGCGCGCGCGCGCGCGCGCTGAGCAAAAGCTGCTAACCCGATTAAGAGC 452
Db 1235 TTTCCGCTGCGCGCGCGCGCGCGCGCTGAGCAAAAGCTGCTAACCCGATTAAGAGC 1294
QY 453 TCCAGACATCTAGCGCCAGGTTTAGCCCGCAGAGACCCGAGCCCGAGGTGAGCATC 512
Db 1295 TCCAGACATCTAGCGCCAGGTTTAGCCCGCAGAGACCCGAGCCCGAGGTGAGCATC 1354
QY 513 CTCCCAACTAGGCGCTCCCACTATCAAGGAGGAGGAGGAGCTCCCTGAGTAC 572
Db 1355 CTCCCAACTAGGCGCTCCCACTATCAAGGAGGAGGAGGAGCTCCCTGAGTAC 1414
QY 573 ATGATGCCCAATTTCAATTTGGGCTCCGCTCACTTAATCCAGAGTACAGGGGCTGGGGT 632
Db 1415 ATGATGCCCAATTTCAATTTGGGCTCCGCTCACTTAATCCAGAGTACAGGGGCTGGGGT 1474
QY 633 CAGGAGAGAGATTTAAAGAACCCACTGTGGGTGAGGGGATAGGGACAGAGAGACATA 692
Db 1475 CAGGAGAGAGATTTAAAGAACCCACTGTGGGTGAGGGGATAGGGACAGAGAGACATA 1534
QY 693 TGGGCAAGCTCTGAGAGAGAGAGAGAGACAAACCTCTGATCTATGAATCTCTGAG 752
Db 1535 TGGGCAAGCTCTGAGAGAGAGAGAGAGAGACAAACCTCTGATCTATGAATCTCTGAG 1594
QY 753 GGCAGAGGAGACAGGAGCTTGAACCTCTTGGCCAGAGGAGAGTGGAGAGAGAGAGA 812
Db 1595 GGCAGAGGAGACAGGAGCTTGAACCTCTTGGCCAGAGGAGAGTGGAGAGAGAGAGA 1654
QY 813 AGGTACAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 872
Db 1655 AGGTACAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1714
QY 873 AAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 932
Db 1715 AAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1774
QY 933 GAGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 992
Db 1775 GAGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1834
QY 993 GGCACATCTGATATAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1052
Db 1835 GGCACATCTGATATAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1094
QY 1053 AACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1112
Db 1895 AACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1954

QY 1113 GGGCTGCTCTTCT 1172
Db 1955 GGGCTGCTCTTCT 2014
QY 1173 ATCAGCTAAGAGAGTGAAGAGCAGGAGCATGAGTGGGCTTGGTCTGAGGTGAGC 1232
Db 2015 ATCAGCTAAGAGAGTGAAGAGCAGGAGCATGAGTGGGCTTGGTCTGAGGTGAGC 2074
QY 1233 CCGAGCTTCTGCGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1291
Db 2075 CCGAGCTTCTGCGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2134
QY 1292 TCTATGATCTATAGGCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1350
Db 2135 TCTATGATCTATAGGCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2194
QY 1351 ACTAGCATTAACCATTTGTTGAGGCGCAATTGAGAGCTTCTCAATCAATCAATCAATC 1410
Db 2195 ACTAGCATTAACCATTTGTTGAGGCGCAATTGAGAGCTTCTCAATCAATCAATCAATC 2254
QY 1411 TCCATTTTCAATTAAGGAGAAATCCCGAGCACTGAGTGTGTGTGTGTGTGTGTGTGT 1470
Db 2255 TCCATTTTCAATTAAGGAGAAATCCCGAGCACTGAGTGTGTGTGTGTGTGTGTGTGT 2314
QY 1471 GGTAGATCTGAACCCAGGAGTGTCAACGCTCTCTCAATC-----CCGACCTTGG 1523
Db 2315 GGTAGATCTGAACCCAGGAGTGTCAACGCTCTCTCAATC-----CCGACCTTGG 2374
QY 1524 GCACTAGAGATTTTCCCTCTCAATTTACTCTCTTAAGCTATGACACCCCTCCGACCTC 1583
Db 2375 GCACTAGAGATTTTCCCTCTCAATTTACTCTCTTAAGCTATGACACCCCTCCGACCTC 2434
QY 1584 TTTCCAGCTGGGAGTGGGAGGAGTCAAGAAAGCCCATCTCCATCTGGATGAGG 1643
Db 2435 TTTCCAGCTGGGAGTGGGAGGAGTCAAGAAAGCCCATCTCCATCTGGATGAGG 2494
QY 1644 ACCCTCATAGCTTTAACCCTGGAGAAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1703
Db 2495 ACCCTCATAGCTTTAACCCTGGAGAAATGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 2554
QY 1704 CTCCCAATTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1763
Db 2555 CTCCCAATTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2614
QY 1764 GGGCAAGTACGCT 1823
Db 2615 GGGCAAGTACGCT 2674
QY 1824 ACTCCCAACACTTGGG--CCCTGGGGAGAGCTGGGAGAGGGGCGGTGGAGAGGCTTG 1882
Db 2675 ACTCCCAACACTTGGG--CCCTGGGGAGAGCTGGGAGAGGGGCGGTGGAGAGGCTTG 2734
QY 1883 ACCCTGAACCTGTATACAAATTAAGAGAGTCTCAGAGA 1923
Db 2735 ACCCTGAACCTGTATACAAATTAAGAGAGTCTCAGAGA 2775

Search completed: October 24, 2004, 11:30:45
Job time : 989 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: October 24, 2004, 11:14:41 (Search time 970 Seconds
(without alignments)
10496.115 Million cell updates/sec

Title: US-09-977-418-7

Perfect score: 1987
Sequence: 1 tgaatgaagcctgcctcatg.....cnnnctactgagagaa 1987

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3407233 seqs, 2561960514 residues

Total number of hits satisfying chosen parameters: 6814465

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09 NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10 NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 20: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
- 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1979	99.6	1987	10	US-09-977-418-7
2	1979	99.6	1987	10	US-09-977-418-7
3	1979	99.6	1987	10	US-09-977-418-7
4	1979	99.6	1987	10	US-09-977-418-7
5	1979	99.6	1987	10	US-09-977-418-7
6	1600.6	83.6	2410	15	US-10-104-047-972
7	414.2	20.8	421	10	US-09-918-995-34469
8	409.4	20.7	412	10	US-09-918-995-5223
9	409.4	20.6	412	10	US-09-918-995-4747
10	402	20.2	402	10	US-09-960-706-32
11	402	20.2	402	10	US-09-873-319-23
12	402	20.2	402	10	US-09-873-367C-532

13	399.4	20.1	486	10	US-09-918-995-3808	Sequence 3808, Ap
14	375	18.9	400	10	US-09-918-995-4147	Sequence 4147, Ap
15	375	18.9	458	9	US-09-864-761-2033	Sequence 2033, Ap
16	277.8	14.0	305	10	US-09-977-418-75	Sequence 75, Ap
17	277.8	14.0	305	10	US-09-977-418-75	Sequence 75, Ap
18	277.8	14.0	305	10	US-09-977-418-75	Sequence 75, Ap
19	277.8	14.0	305	10	US-09-977-418-75	Sequence 75, Ap
20	277.8	14.0	305	10	US-09-977-418-75	Sequence 75, Ap
21	277.8	14.0	305	10	US-09-977-418-75	Sequence 75, Ap
22	171.4	8.6	719	15	US-10-037-270-669	Sequence 669, Ap
23	154	7.8	472	15	US-09-918-995-5152	Sequence 5152, Ap
24	147	7.4	522	15	US-10-029-186-1250	Sequence 1250, A
25	105.6	5.3	505	10	US-09-918-995-1046	Sequence 1046, Ap
26	105	5.3	134	10	US-09-918-995-1166	Sequence 1166, Ap
27	105	5.3	411	10	US-09-918-995-1761	Sequence 1761, Ap
28	105	5.3	480	10	US-09-918-995-4429	Sequence 4429, Ap
29	105	5.3	482	10	US-09-918-995-5152	Sequence 5152, Ap
30	104	5.2	473	10	US-09-918-995-5152	Sequence 5152, Ap
31	100.6	5.1	462	10	US-09-918-995-5501	Sequence 5501, Ap
32	100.6	5.1	464	10	US-09-918-995-5196	Sequence 5196, Ap
33	85	4.3	196	15	US-10-029-186-1250	Sequence 1250, A
34	75	3.8	192	9	US-09-864-761-1878	Sequence 1878, A
35	60	3.0	60	10	US-09-908-975-11540	Sequence 11540, A
36	60	3.0	60	10	US-09-908-975-11755	Sequence 11755, A
37	60	3.0	60	10	US-09-908-975-11892	Sequence 11892, A
38	60	3.0	60	10	US-09-908-975-11893	Sequence 11893, A
39	54.4	2.7	65	10	US-09-908-975-13947	Sequence 13947, A
40	43	2.2	906	15	US-10-369-493-12907	Sequence 12907, A
41	42.2	2.1	630	17	US-10-437-963-18774	Sequence 18774, A
42	41.8	2.1	674	16	US-10-424-599-69089	Sequence 69089, A
43	41.6	2.1	1049	14	US-10-123-155-358	Sequence 358, Ap
44	41.6	2.1	1049	15	US-10-146-731-358	Sequence 358, Ap
45	41.6	2.1	1449	15	US-10-140-472-358	Sequence 358, Ap

ALIGNMENTS

RESULT 1
US-09-977-418-7
Sequence 7, Application US/09977418
Publication No. US20030027158A1 polynucleotides and polypeptides encoded t
GENERAL INFORMATION:
APPLICANT: Shinketsu et al
TITLE OF INVENTION: No. US20030027158A1 polynucleotides and polypeptides encoded t
FILE REFERENCE: 15966-552
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: USSN 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: USSN 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: USSN 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: USSN 60/201,388
PRIOR FILING DATE: 2000-05-03
NUMBER OF SEQ ID NOS: 93
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 1987
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (991)..(1446)
NAME/KEY: Variation
LOCATION: (1)..(1981)
OTHER INFORMATION: where n can be any nucleotide
US-09-977-418-7
Query Match 99.6%; Score 1979; DB 10; Length 1987;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1987; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAATGAGGCTGTCCATGCTGTGTAAGAGCCCATCCCAAGAGCCCGGTGGCA 60
Db 1 TGAATGAGGCTGTCCATGCTGTGTAAGAGCCCATCCCAAGAGCCCGGTGGCA 60

QY 61 GGCAACTCAAGACCTTAGAGAGCGCTTAGAGTTGCGGTGAGCTGAGAGACTTTCTCA 120
Db 61 GGCAACTCAAGACCTTAGAGAGCGCTTAGAGTTGCGGTGAGCTGAGAGACTTTCTCA 120

QY 121 CTTGAAGACATCATTTGTGACCACTCCCAACCAACCAATCGAGGTGGGGGCTGAGA 180
Db 121 CTTGAAGACATCATTTGTGACCACTCCCAACCAACCAATCGAGGTGGGGGCTGAGA 180

QY 181 AGCTGGCGCTGAAACGGAACATGTATCAATGAAACCTTACGACTTACAAAGTGCAGCT 240
Db 181 AGCTGGCGCTGAAACGGAACATGTATCAATGAAACCTTACGACTTACAAAGTGCAGCT 240

QY 241 GCCGAGAGAGCTGAGACCCCGAGGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 300
Db 241 GCCGAGAGAGCTGAGACCCCGAGGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 300

QY 301 CTTCACTAGTCCGGGAGCGGTGACCCGGAACAGAAACGTCAGAGAGCTTCCGAGC 360
Db 301 CTTCACTAGTCCGGGAGCGGTGACCCGGAACAGAAACGTCAGAGAGCTTCCGAGC 360

QY 361 GGAATCAAAATGTAGTGTCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 420
Db 361 GGAATCAAAATGTAGTGTCTCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 420

QY 421 CAGCAAAAGCTCGCTAACCCATTACAGAGCTCCAGAGCATCTAGCCAGCTTCTAGC 480
Db 421 CAGCAAAAGCTCGCTAACCCATTACAGAGCTCCAGAGCATCTAGCCAGCTTCTAGC 480

QY 481 CCCCAGGACCCCAAGCCCAAGGTGACATCTCTCCCAACCTAGAGGCTTCCACTTATC 540
Db 481 CCCCAGGACCCCAAGCCCAAGGTGACATCTCTCCCAACCTAGAGGCTTCCACTTATC 540

QY 541 CAGGAGAGGCGAGGAGCTCCCTGCTGACATGATGCGGAGTTTGGATTGGCTTC 600
Db 541 CAGGAGAGGCGAGGAGCTCCCTGCTGACATGATGCGGAGTTTGGATTGGCTTC 600

QY 601 CGTCACTTATTCAGAGTACAGAGGCTGGGGTCCAGGAAAGAACTTAAAGACCACT 660
Db 601 CGTCACTTATTCAGAGTACAGAGGCTGGGGTCCAGGAAAGAACTTAAAGACCACT 660

QY 661 GTGGGTCAAGGGAATGGACACAGAGACATATGGGCAAGCTCTGACAGGACAGACAGA 720
Db 661 GTGGGTCAAGGGAATGGACACAGAGACATATGGGCAAGCTCTGACAGGACAGACAGA 720

QY 721 GACAAACCTCTGATCTATGAAAGTCTTGCAGAGGACAGGAGACCTGGAACCTT 780
Db 721 GACAAACCTCTGATCTATGAAAGTCTTGCAGAGGACAGGAGACCTGGAACCTT 780

QY 781 CTTGGCAAGGGGAGTGGAGAGACAGAGGGAAGGTCAAGGCAAGGCTGCTATCTAG 840
Db 781 CTTGGCAAGGGGAGTGGAGAGACAGAGGGAAGGTCAAGGCAAGGCTGCTATCTAG 840

QY 841 TGAATCTAATTCGCCAGGGCTCAGCAAGGCCAAGAGAGACAGCCCTGACGCTAACTT 900
Db 841 TGAATCTAATTCGCCAGGGCTCAGCAAGGCCAAGAGAGACAGCCCTGACGCTAACTT 900

QY 901 CCCCCTTACAGCTTCAGAGCCCAAGCCAGGAGGAGGAGCTGCTGCCACCCCTGCTCC 960
Db 901 CCCCCTTACAGCTTCAGAGCCCAAGCCAGGAGGAGGAGCTGCTGCCACCCCTGCTCC 960

QY 961 CCAGCAGCTGCTGTGCAAGGGGAGAGCATGCCATCTGATATGATGGGTTT 1020
Db 961 CCAGCAGCTGCTGTGCAAGGGGAGAGCATGCCATCTGATATGATGGGTTT 1020

QY 1021 CCAATACAGCTGCTGTGAAAAAAGTCAATGAACTCTGCGCTTCCGCTGCTGCGG 1080
Db 1021 CCAATACAGCTGCTGTGAAAAAAGTCAATGAACTCTGCGCTTCCGCTGCTGCGG 1080

Db 1021 CCAATACAGCTGCTGTGAAAAAAGTCAATGAACTCTGCGCTTCCGCTGCTGCGG 1080

QY 1081 CCTCAGGCAAGGCAAGTGGGGTGGGGGGTGGGCTTCTCTCCCTCCCAAGAGCC 1140
Db 1081 CCTCAGGCAAGGCAAGTGGGGTGGGGGGTGGGCTTCTCTCCCTCCCAAGAGCC 1140

QY 1141 TGTGTTCTGGGGCTGCTCCCATCAGACAGATCACTTAAAGAGTGAAGCAGGCT 1200
Db 1141 TGTGTTCTGGGGCTGCTCCCATCAGACAGATCACTTAAAGAGTGAAGCAGGCT 1200

QY 1201 ATGATGAGGCTTGGGCTCCAGAGTGGAGCCCACTTGGCCACTTCCCTCCG 1260
Db 1201 ATGATGAGGCTTGGGCTCCAGAGTGGAGCCCACTTGGCCACTTCCCTCCG 1260

QY 1261 CAGTACAGCTCCATCCATCCCTCTTATCATGAACTATAGGCTGCTGTGTGA 1320
Db 1261 CAGTACAGCTCCATCCATCCCTCTTATCATGAACTATAGGCTGCTGTGTGA 1320

QY 1321 ACAACACACCCCTATGCTGCTCCCAATCAATCAGATTAACATGTTGAGGCCAAT 1380
Db 1321 ACAACACACCCCTATGCTGCTCCCAATCAATCAGATTAACATGTTGAGGCCAAT 1380

QY 1381 TCAGAGCTTCTCAATCAATCAATTTCAATCTCCATTTTCAATCGGGGAAACATCCCG 1440
Db 1381 TCAGAGCTTCTCAATCAATCAATTTCAATCTCCATTTTCAATCGGGGAAACATCCCG 1440

QY 1441 AGCCAGTGAAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1500
Db 1441 AGCCAGTGAAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1500

QY 1501 TGCTCTCACTCCCACTCTGAGGCACTGAGAGTATTTCCCTCATCTCACTCTTCA 1560
Db 1501 TGCTCTCACTCCCACTCTGAGGCACTGAGAGTATTTCCCTCATCTCACTCTTCA 1560

QY 1561 GGTATGACACCCCTCCCAAGTCTTCAAGTGGGGAGTCAATGAGAAAGCC 1620
Db 1561 GGTATGACACCCCTCCCAAGTCTTCAAGTGGGGAGTCAATGAGAAAGCC 1620

QY 1621 CCAATCTCCATCTGGGATAGGAGCTTCCATCAGGCTTAAAGCTGGGAAATGCTGCTG 1680
Db 1621 CCAATCTCCATCTGGGATAGGAGCTTCCATCAGGCTTAAAGCTGGGAAATGCTGCTG 1680

QY 1681 CCCCAGTGACTCTTGGTTGCTCTCCCAATACAGAGGAGTGGAGGAGGAGTGG 1740
Db 1681 CCCCAGTGACTCTTGGTTGCTCTCCCAATACAGAGGAGTGGAGGAGGAGTGG 1740

QY 1741 GTCTCAGTTAGCAAGGGGCTCCCAAGGCAAGTCAAGCTCTCCATGCTCTCTGCTC 1800
Db 1741 GTCTCAGTTAGCAAGGGGCTCCCAAGGCAAGTCAAGCTCTCCATGCTCTCTGCTC 1800

QY 1801 AGTGTGCTTAAAGGTGCTCTCACTCCCAAGCTCTGAGGCTTGGGGAGGAGCTGGGG 1860
Db 1801 AGTGTGCTTAAAGGTGCTCTCACTCCCAAGCTCTGAGGCTTGGGGAGGAGCTGGGG 1860

QY 1861 AGGGGCTGAGGAGGAGGCTTGAAGCTTGAATCAATTAAGAGACAGTCTCAC 1920
Db 1861 AGGGGCTGAGGAGGAGGCTTGAAGCTTGAATCAATTAAGAGACAGTCTCAC 1920

QY 1921 AGACNTCTGAGGAGGCTGCTGCGAGAGTCTCAATCTTGGGAGGAGGAGGAGGAGGAG 1980
Db 1921 AGACNTCTGAGGAGGCTGCTGCGAGAGTCTCAATCTTGGGAGGAGGAGGAGGAGGAG 1980

QY 1981 AGAGAAA 1987
Db 1981 AGAGAAA 1987

RESULT 2
US-09-977-033A-7
; Sequence 7, Application US/09977033A
; Publication No. US20030082554A1
; GENERAL INFORMATION:
; APPLICANT: Shimkels, Richard A

QY 1441 AGCCACTGAGTGTGCTGCTTTGTGTCATGGAAGTTAGTGTGAACCCAGGGGTCTCAACNGC 1500
DB 1441 AGCCACTGAGTGTGCTGCTTTGTGTCATGGAAGTTAGTGTGAACCCAGGGGTCTCAACNGC 1500
QY 1501 TGTCTCAACTCCCACTGCTGGGACCTGAGAGTATTTCCCTCAATTTCACTCTCTTA 1560
DB 1501 TGTCTCAACTCCCACTGCTGGGACCTGAGAGTATTTCCCTCAATTTCACTCTCTTA 1560
QY 1561 GGTATGACCCCTCCCACTGCTTCCAGCTGGGGGATGGGGGAGTCAATAGAAAAGCC 1620
DB 1561 GGTATGACCCCTCCCACTGCTTCCAGCTGGGGGATGGGGGAGTCAATAGAAAAGCC 1620
QY 1621 CCATCTCCCATCTGGGATAGGAGACCTTCCATCAGCCTTAACTCTGGAAATCTCTGTG 1680
DB 1621 CCATCTCCCATCTGGGATAGGAGACCTTCCATCAGCCTTAACTCTGGAAATCTCTGTG 1680
QY 1681 CCCCAGTGACTCTTGGTTTGTCTCCCATACAGAGAGGAGTGAAGGAGAGGAGTGG 1740
DB 1681 CCCCAGTGACTCTTGGTTTGTCTCCCATACAGAGAGGAGTGAAGGAGAGGAGTGG 1740
QY 1741 GTCTCAATTAGCAGGGGTCCCAAGGCAAGTCAAGCTCTCTCCATGCTCTCTGTGTC 1800
DB 1741 GTCTCAATTAGCAGGGGTCCCAAGGCAAGTCAAGCTCTCTCCATGCTCTCTGTGTC 1800
QY 1801 AGTGTGCTTAGGGGTGGCTCTCTCACTCCCACTCTGGGGCCCTTGGGGGAGAGTGGGG 1860
DB 1801 AGTGTGCTTAGGGGTGGCTCTCTCACTCCCACTCTGGGGCCCTTGGGGGAGAGTGGGG 1860
QY 1861 AGGGGGCGGTGGAGAGCCCTGACGCTGGAACCTGTATAACAATTAAGAGACAGTCTAC 1920
DB 1861 AGGGGGCGGTGGAGAGCCCTGACGCTGGAACCTGTATAACAATTAAGAGACAGTCTAC 1920
QY 1921 AGACNTCTGAGGCGCGCTGCGGAGTCTCAAACTTGGGGCAGGCGNNNCTTAATTG 1980
DB 1921 AGACNTCTGAGGCGCGCTGCGGAGTCTCAAACTTGGGGCAGGCGNNNCTTAATTG 1980
QY 1981 AGAGAAA 1987
DB 1981 AGAGAAA 1987

RESULT 3

US-09-977-751C-7
Sequence 7, Application US/0977751C
Publication No. US20030134430A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Hernandez, Elma
APPLICANT: Herrman, John
TITLE OF INVENTION: No. US20030134430A1 Amino Acid Sequences for Human Caenorhabditis
FILE REFERENCE: 15966-552 CON S-40
CURRENT APPLICATION NUMBER: US/09/977,751C
PRIOR FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 1987
TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (991)..(1446)
FEATURE:
NAME/KEY: Misc Feature
LOCATION: (1148)..(1148)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:
NAME/KEY: Misc Feature
LOCATION: (1498)..(1498)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:
NAME/KEY: Misc Feature
LOCATION: (1925)..(1925)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:
NAME/KEY: Misc Feature
LOCATION: (1943)..(1943)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:
NAME/KEY: Misc Feature
LOCATION: (1965)..(1972)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:
NAME/KEY: Variant
LOCATION: (53)..(53)
OTHER INFORMATION: wherein Xaa is any amino acid as defined in the specification.

US-09-977-751C-7

Query Match 99.6%; Score 1979; DB 10; Length 1987;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1987; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAATGAAGGCTGTTCATGCTCTGTCATGAAGCCATCCCAAGCCCGGTGGGA 60
DB 1 TGAATGAAGGCTGTTCATGCTCTGTCATGAAGCCATCCCAAGCCCGGTGGGA 60
QY 61 GGCAACATCAAGCCCTAGAGAGCGCTATGAGTTGGGTGACGTGAGCTTCTCA 120
DB 61 GGCAACATCAAGCCCTAGAGAGCGCTATGAGTTGGGTGACGTGAGCTTCTCA 120
QY 121 CTTGAAGACATCAATTTGACACCACTTCCAAACCAACATCAGGTGGGGGTAGAA 180
DB 121 CTTGAAGACATCAATTTGACACCACTTCCAAACCAACATCAGGTGGGGGTAGAA 180
QY 181 AGCTGGGGGTGAAAGGACACTGTATGAAACCTTACGACTACACAAGTCCAGACT 240
DB 181 AGCTGGGGGTGAAAGGACACTGTATGAAACCTTACGACTACACAAGTCCAGACT 240
QY 241 GCGGAGAGAGCTGAGACCCGAGCGGTGACCTGCGCTACCTGTGCGGAGAGCGGAC 300
DB 241 GCGGAGAGAGCTGAGACCCGAGCGGTGACCTGCGCTACCTGTGCGGAGAGCGGAC 300
QY 301 CTTCACTAGTCCGGGACAGCGGTACCCGATACAGAACGTCAGACACCTTCGGAC 360
DB 301 CTTCACTAGTCCGGGACAGCGGTACCCGATACAGAACGTCAGACACCTTCGGAC 360
QY 361 GGAGATCAAAATGAGTGGCTCTCCCTCCCTTCCCTTCCCTGCGGCGGCGGACGCT 420
DB 361 GGAGATCAAAATGAGTGGCTCTCCCTCCCTTCCCTTCCCTGCGGCGGCGGACGCT 420
QY 421 CAGCAAGGCTGCTAACCCTTACCAAGCTTCCAGAGCATCTGACGCTTCTAGC 480
DB 421 CAGCAAGGCTGCTAACCCTTACCAAGCTTCCAGAGCATCTGACGCTTCTAGC 480
QY 481 CCCCAGGACCCCAAGCCCAAGGTGAGACATCTCCCAACCTAGAGGCTCTCACTATC 540
DB 481 CCCCAGGACCCCAAGCCCAAGGTGAGACATCTCCCAACCTAGAGGCTCTCACTATC 540
QY 541 CAGGGGAGGCAAGGACTCTGCTGCTGACATGATGCGGATTTGAGATTGGCTTC 600
DB 541 CAGGGGAGGCAAGGACTCTGCTGCTGACATGATGCGGATTTGAGATTGGCTTC 600

/ NAME/KEY: Misc_Feature
 / LOCATION: (1943)...(1943)
 / OTHER INFORMATION: wherein n is a or c or t or g
 / FEATURE:
 / NAME/KEY: Misc_Feature
 / LOCATION: (1969)...(1972)
 / OTHER INFORMATION: wherein n is a or c or t or g
 / FEATURE:
 / NAME/KEY: Variant
 / LOCATION: (53)...(53)
 / OTHER INFORMATION: wherein Xaa is any amino acid as defined in the specification.
 / US-09-977-639A-7

Query Match 99.6%; Score 1979; DB 10; Length 1987;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1987; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAATGAAGGCTGTCCATCTCTGTCAAGAGCCCATCCCAACCAAGCCCGGTGGGA 60
 DB 1 TGAATGAAGGCTGTCCATCTCTGTCAAGAGCCCATCCCAACCAAGCCCGGTGGGA 60
 QY 61 GGCAACTCAAGACCTTAGAGAGCCCTATGAGTTTGGGTGACGTGAGACTTTCTCA 120
 DB 61 GGCAACTCAAGACCTTAGAGAGCCCTATGAGTTTGGGTGACGTGAGACTTTCTCA 120
 QY 121 CTTGAAGACATCTATTGTACACCTCCAAACAAACCAATCGAGGTGGGGCTGAGAA 180
 DB 121 CTTGAAGACATCTATTGTACACCTCCAAACAAACCAATCGAGGTGGGGCTGAGAA 180
 QY 181 AGCTGGCGCTGAACCGACACTGTATACATGAAACACTTACAGACTACAAAGTCCAGACT 240
 DB 181 AGCTGGCGCTGAACCGACACTGTATACATGAAACACTTACAGACTACAAAGTCCAGACT 240
 QY 241 GCGGAGAGAGCTGAGACCCCGAGGTGGGTGACCTGCTGCTAGTGGGAGAGAGAGACAG 300
 DB 241 GCGGAGAGAGCTGAGACCCCGAGGTGGGTGACCTGCTGCTAGTGGGAGAGAGAGACAG 300
 QY 301 CCTCACTAGTCCCGGACCGGTGACCCCGATACAGAAACAGTCCAGACAGACTTCCGAGAC 360
 DB 301 CCTCACTAGTCCCGGACCGGTGACCCCGATACAGAAACAGTCCAGACAGACTTCCGAGAC 360
 QY 361 GGAGATCAAAATTTGAGTGTCTCTCCCTTCCCTTCCCTGTCGCCGCCGCCACGCTGCG 420
 DB 361 GGAGATCAAAATTTGAGTGTCTCTCCCTTCCCTTCCCTGTCGCCGCCGCCACGCTGCG 420
 QY 421 CAGCAAGAGCTCGCTAACCCATTACACAGCTCCAGACATCTCAGCCAGGTTCTAGC 480
 DB 421 CAGCAAGAGCTCGCTAACCCATTACACAGCTCCAGACATCTCAGCCAGGTTCTAGC 480
 QY 481 CCCACGCAACCCGACCCGAGGTGACATCTCCCAAATGAGGCTCCACTCTATC 540
 DB 481 CCCACGCAACCCGACCCGAGGTGACATCTCCCAAATGAGGCTCCACTCTATC 540
 QY 541 CAGGAGAGGAGGAGGAGCTCCCTGGCTGACATGATGCCAGATTTTCAAGTTTGGCTC 600
 DB 541 CAGGAGAGGAGGAGGAGCTCCCTGGCTGACATGATGCCAGATTTTCAAGTTTGGCTC 600
 QY 601 CGTCACTTATTCAGAGTACAGGAGCTGAGGCTCAGGAGAGAGATCTTAAAGAACCACT 660
 DB 601 CGTCACTTATTCAGAGTACAGGAGCTGAGGCTCAGGAGAGAGATCTTAAAGAACCACT 660
 QY 661 GTGGGTCAAGGGAATGGGACAGAGATATATGGGCAAGCTCTGACAGACAGAGACA 720
 DB 661 GTGGGTCAAGGGAATGGGACAGAGATATATGGGCAAGCTCTGACAGACAGAGACA 720
 QY 721 GACAAACCTCTGATCTATGAGTGTCTGACAGGGAAGGGAACAGGGAACCTGAAACCT 780
 DB 721 GACAAACCTCTGATCTATGAGTGTCTGACAGGGAAGGGAACAGGGAACCTGAAACCT 780
 QY 781 CTTGGGCAAGGGAAGTGGAGAGACAGAGGAAAGGTCAAGGCAAGGCTGCTATCTAAG 840
 DB 781 CTTGGGCAAGGGAAGTGGAGAGACAGAGGAAAGGTCAAGGCAAGGCTGCTATCTAAG 840

QY 841 TGAATCTAATTGCCGAGGCTCAGACAGGCCAAGAGAGACAGCCGTGACGTAACCT 900
 DB 841 TGAATCTAATTGCCGAGGCTCAGACAGGCCAAGAGAGACAGCCGTGACGTAACCT 900
 QY 901 CCCCTCTACAGGCTCCCAAGCCCAACGACGACGACGAGCTGCTGCCACCCGTCGCC 960
 DB 901 CCCCTCTACAGGCTCCCAAGCCCAACGACGACGACGAGCTGCTGCCACCCGTCGCC 960
 QY 961 CCAGCAGCTGCTGTGTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 DB 961 CCAGCAGCTGCTGTGTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 QY 1021 CCAATACAGCTGTGTTCTGAAAAAAGTGAATGAATCTCTGCGCTGCGCTGCTGCGG 1080
 DB 1021 CCAATACAGCTGTGTTCTGAAAAAAGTGAATGAATCTCTGCGCTGCGCTGCTGCGG 1080
 QY 1081 CCTCAGGCAAGGCAAGTGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1140
 DB 1081 CCTCAGGCAAGGCAAGTGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1140
 QY 1141 TGTGTTCTGAGGCTGTGCTCCATGACAGAGATGACCTTAACAGAGTGAAGCCAGGC 1200
 DB 1141 TGTGTTCTGAGGCTGTGCTCCATGACAGAGATGACCTTAACAGAGTGAAGCCAGGC 1200
 QY 1201 ATGATGGGGCTTTGGGCTCCGAGGTTGAGACCCAGCTCTGCAACCTTCCCTCCGG 1260
 DB 1201 ATGATGGGGCTTTGGGCTCCGAGGTTGAGACCCAGCTCTGCAACCTTCCCTCCGG 1260
 QY 1261 CAGTCACTCTCCATCCCTCTTATCTATGATGATCTATGAGCTGCTGCTGCTGCT 1320
 DB 1261 CAGTCACTCTCCATCCCTCTTATCTATGATGATCTATGAGCTGCTGCTGCTGCTGCT 1320
 QY 1321 ACAACAGACCCCTATCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
 DB 1321 ACAACAGACCCCTATCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
 QY 1381 TCAGAGCTTCTCAATACAGATTTACATCTCATTTTCAATTAACGGGGAACATCCCG 1440
 DB 1381 TCAGAGCTTCTCAATACAGATTTACATCTCATTTTCAATTAACGGGGAACATCCCG 1440
 QY 1441 AGCCATGAGTGTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 DB 1441 AGCCATGAGTGTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 QY 1501 TGCTCTCAATCCGCACTGTGGGACAGTGAAGTATTTCCCTCACTTCTACCTCTCTA 1560
 DB 1501 TGCTCTCAATCCGCACTGTGGGACAGTGAAGTATTTCCCTCACTTCTACCTCTCTA 1560
 QY 1561 GGTATGACACCTTCCGACGTCTTCAAGTGGGAGTGGGAGTCAATAGGAAAGCC 1620
 DB 1561 GGTATGACACCTTCCGACGTCTTCAAGTGGGAGTGGGAGTCAATAGGAAAGCC 1620
 QY 1621 CCCATCTCCATCTGGGATGGGACCTTCCATGAGCTTAACTTGGGAAATGCTGCTG 1680
 DB 1621 CCCATCTCCATCTGGGATGGGACCTTCCATGAGCTTAACTTGGGAAATGCTGCTG 1680
 QY 1681 CCCCACTGACTCTTGTGTTCTGCTCCCATACATGAGAGAGGAGTGAAGGAGTGG 1740
 DB 1681 CCCCACTGACTCTTGTGTTCTGCTCCCATACATGAGAGAGGAGTGAAGGAGTGG 1740
 QY 1741 GTCTCAGTTAGCAGGAGTCCGAGGAGAGTCAAGCTCTCTCCATGCTCTCTGCTG 1800
 DB 1741 GTCTCAGTTAGCAGGAGTCCGAGGAGAGTCAAGCTCTCTCCATGCTCTCTGCTG 1800
 QY 1801 AGTGTGCTTAAAGGAGTGGCTCTCACTCCACATCTCTGAGGCTTGGGAGAGAGCTG 1860
 DB 1801 AGTGTGCTTAAAGGAGTGGCTCTCACTCCACATCTCTGAGGCTTGGGAGAGAGCTG 1860
 QY 1861 AGGAGGCTGAGGAGAGCTCTGACGCTGGAACCTGATACAAATTAAGAGACAGTCTAC 1920
 DB 1861 AGGAGGCTGAGGAGAGCTCTGACGCTGGAACCTGATACAAATTAAGAGACAGTCTAC 1920
 QY 1921 AGACNTCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
 DB 1921 AGACNTCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980

QY 1081 CTTCCAGGCAAGGCGAAGTGGGGTGGGGGCTGGTCTTCTCCCTCCCAAGGCC 1140
 Db 1081 CTTCCAGGCAAGGCGAAGTGGGGTGGGGGCTGGTCTTCTCCCTCCCAAGGCC 1140
 QY 1141 TGTGTTTGGGGTGTCTCCATGTCAGACAGATCACTTAACAGAGATGAAAGCCAGGC 1200
 Db 1141 TGTGTTTGGGGTGTCTCCATGTCAGACAGATCACTTAACAGAGATGAAAGCCAGGC 1200
 QY 1201 ATGAGTGGGGCTTTGGGCTCTGAGGTGAGCCCAAGCTTTGCGACCTTCCCTCCG 1260
 Db 1201 ATGAGTGGGGCTTTGGGCTCTGAGGTGAGCCCAAGCTTTGCGACCTTCCCTCCG 1260
 QY 1261 CAGTCAAGCTCTCATCATCCCTCTCTTAATCTAATGAACTATAGGCTCGGTGTGA 1320
 Db 1261 CAGTCAAGCTCTCATCATCCCTCTCTTAATCTAATGAACTATAGGCTCGGTGTGA 1320
 QY 1321 ACAACACACCCCTATGTTGTCTTCAAACTACAGATTAACATTTGTTGAGCCAAAT 1380
 Db 1321 ACAACACACCCCTATGTTGTCTTCAAACTACAGATTAACATTTGTTGAGCCAAAT 1380
 QY 1381 TCGAGCTTCTCAAACTACAGATTAACATTTGTTGAGCCAAAT 1440
 Db 1381 TCGAGCTTCTCAAACTACAGATTAACATTTGTTGAGCCAAAT 1440
 QY 1441 AGCCACTGAGTGTGTGTCTTGTCACTGAGGTGAATCTGAACCCAGGCTGTCAACGC 1500
 Db 1441 AGCCACTGAGTGTGTGTCTTGTCACTGAGGTGAATCTGAACCCAGGCTGTCAACGC 1500
 QY 1501 TGTCTTCAACTCTCCCACTCTTGGGCACTGAGAGATTTTCCCTCACTTCACTCTTA 1560
 Db 1501 TGTCTTCAACTCTCCCACTCTTGGGCACTGAGAGATTTTCCCTCACTTCACTCTTA 1560
 QY 1561 GGTATGACACCCCTCCCAAGCTCTTCCAGCTGGGGATGGGGGATCTATAGAAAAGCC 1620
 Db 1561 GGTATGACACCCCTCCCAAGCTCTTCCAGCTGGGGATGGGGGATCTATAGAAAAGCC 1620
 QY 1621 CCATCTCCCATCTGGGATAGGAGCTTCCATGAGCTTAACTCCCTGGGAAATGCTCTG 1680
 Db 1621 CCATCTCCCATCTGGGATAGGAGCTTCCATGAGCTTAACTCCCTGGGAAATGCTCTG 1680
 QY 1681 CCCCCAGTACTCTTGGTTTGTCTCCCACTGACAGAGAGGATGAGGAGGAGAGGATG 1740
 Db 1681 CCCCCAGTACTCTTGGTTTGTCTCCCACTGACAGAGAGGATGAGGAGGAGAGGATG 1740
 QY 1741 GTCTCAGTTAGAGAGGAGTCCCAAGGCAAGTCAAGCTCTCCCTCCATGCTCTGCTC 1800
 Db 1741 GTCTCAGTTAGAGAGGAGTCCCAAGGCAAGTCAAGCTCTCCCTCCATGCTCTGCTC 1800
 QY 1801 AGTGTGCTTGAAGGAGTCTCTCACTCCCACTGAGGCTCTGAGGAGGAGAGTGGGG 1860
 Db 1801 AGTGTGCTTGAAGGAGTCTCTCACTCCCACTGAGGCTCTGAGGAGGAGAGTGGGG 1860
 QY 1861 AGGGGGCCGTGGAGAGAGCCCTGACGCTGAACTGTATACAAATTAAGACAGTCTAC 1920
 Db 1861 AGGGGGCCGTGGAGAGAGCCCTGACGCTGAACTGTATACAAATTAAGACAGTCTAC 1920
 QY 1921 AGACNTCTGAGAGGCGCCCTGCGCNGAGTTCTCAACTTAAGGAGGCGNNNCTTAATTG 1980
 Db 1921 AGACNTCTGAGAGGCGCCCTGCGCNGAGTTCTCAACTTAAGGAGGCGNNNCTTAATTG 1980
 QY 1981 AGAGAAA 1987
 Db 1981 AGAGAAA 1987

RESULT 6
 US-10-104-047-972
 ; Sequence 972, Application US/10104047
 ; Publication No. US20030236392A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HELIX RESEARCH INSTITUTE
 ; TITLE OF INVENTION: No. US20030236392A1 full length cDNA
 ; FILE REFERENCE: H1-A0105

QY 33 GCCCATCCCAACCAAGCCCGGTGGGCGAGGCAATCAAGACCTTAGAGAGCGCTATGA 92
 Db 539 GGCCTTCCCAAGCCCGCCCGGTGGGCGAGGCAATCAAGACCTTAGAGAGCGCTATGA 598
 QY 93 GTTGGCGGTGAGCGTGAAGATCTTCACTTGAAGATATATGTCAACCACTTCAACCA 152
 Db 599 GTTGGCGGTGAGCGTGAAGATCTTCACTTGAAGATATATGTCAACCACTTCAACCA 152
 QY 153 AACCAATGAGGTGGGCGGTGAAGAACTGGCGGCTGAACGCACTGTATCATGAAA 212
 Db 656 CAACCAATGAGGTGGGCGGTGAAGAACTGGCGGCTGAACGCACTGTATCATGAAA 212
 QY 213 CACCTTACGATACACAGTGTCCAGATGTCCCGAGACGTGAGACCCCGAGCGGTGA 272
 Db 714 TTGCT 762
 QY 273 CCTGGGCTACTGTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 332
 Db 763 CTTGGG---CTTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 819
 QY 333 CAGAACAGTGCAGACAGCTTCCGAGCGAGATCAAAATCTAGTGTCTCTCTCTCTCT 392
 Db 820 CAGAACAGTGCAGACAGCTTCCGAGCGAGATCAAAATCTAGTGTCTCTCTCTCTCT 392
 QY 393 TTTCCCTGCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 452
 Db 880 TTTCCCTGCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 939
 QY 453 TCCAGACATCTGAGCCAGGATCTAGCCCGCCAGGAGGAGGAGGAGGAGGAGGAGG 512
 Db 940 TCCAGACATCTGAGCCAGGATCTAGCCCGCCAGGAGGAGGAGGAGGAGGAGGAGG 939
 QY 513 CTCCCAACTAGGAGGCTCTCACTCTAATCCAGGAGGAGGAGGAGGAGGAGGAGGAG 572
 Db 1000 CTCCCAACTAGGAGGCTCTCACTCTAATCCAGGAGGAGGAGGAGGAGGAGGAGGAG 1059
 QY 573 ATGATGCCAGATTTTCAAGTTTGGGCTCTCTCACTTAACTCAAGATTAAGGAGGAG 632
 Db 1060 ATGATGCCAGATTTTCAAGTTTGGGCTCTCTCACTTAACTCAAGATTAAGGAGGAG 1119
 QY 633 CAGGAAAGAAATCTTAAGAACCCACTGTGGGTCAAGGGAATGGGAGCAGAGACATA 692
 Db 1120 CAGGAAAGAAATCTTAAGAACCCACTGTGGGTCAAGGGAATGGGAGCAGAGACATA 1179
 QY 693 TGGGCAAGCTCTGAGAGGAGCAGACAGACAGACAAACCTCTGATCTATGAAGTTCTGAG 752
 Db 1180 TGGGCAAGCTCTGAGAGGAGCAGACAGACAGACAAACCTCTGATCTATGAAGTTCTGAG 1239
 QY 753 GGCAGAGGAGCAGAGGAGCCTGGAACCTCTTGGCCAAAGGAGGTGGAAGAGAGAGGA 812
 Db 1240 GGCAGAGGAGCAGAGGAGCCTGGAACCTCTTGGCCAAAGGAGGTGGAAGAGAGAGGA 1299
 QY 813 AGGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 872
 Db 1300 AGGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1359
 QY 873 AAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 932

Query Match 83.6%; Score 1660.6; DB 15; Length 2410;
 Best Local Similarity: 95.6%; Pred. No. 0;
 Matches 1808; Conservative 0; Mismatches 61; Indels 22; Gaps 9;

Db	1360	AAGAGGAGACAGCCGTGAAGGTAAACTTCCCTCTCAACGAGCTCCAAAGCCCCAGCCAGC	141.9
QY	933	GAGCAGAGCTGCTGCTCCCAACCCCCGTGCCCCCAGCCAGCTGAGCTGTGCCAGGGCAGAGCCAT	992
Db	1420	GAGAGAGGTGCTGCTCCCAACCCCCGTGCCCCCAGCCAGCTGAGCTGTGCCAGGGCAGAGCCAT	147.9
QY	993	GCCACATCTGTATATAGATGGGGTTTTTCCATACAGCTGGTGTGCGTAAATACTGATGA	1052.2
Db	1480	GCCACATCTGTATATAGATGGGGTTTTTCCATACAGCTGGTGTGCGTAAATACTGATGA	153.9
QY	1053	AACCTCTGCGCTCTGCGCGCTGTGCTGGGGCTTCAGGCAAGGCCAAATGTGGGTTGGGGGTG	111.2
Db	1540	AACCTCTGCGCTCTGCGCGCTGTGCTGGGGCTTCAGGCAAGGCCAAATGTGGGTTGGGGGTG	159.9
QY	1113	GGGCTGTGCTCTTCCCTCCCAAGAGGCTGTGTTGTTGTTGGGGCTGTGCCATGACAGACAG	117.2
Db	1600	GGGCTGTGCTCTTCCCTCCCAAGAGGCTGTGTTGTTGTTGGGGCTGTGCCATGACAGACAG	165.9
QY	1173	ATCACCTTAACAGAGATGGAGCCAGGGCATGAGATGGGCTTTGGGTCTGAGGTTGAC	123.2
Db	1660	ATCACCTTAACAGAGATGGAGCCAGGGCATGAGATGGGCTTTGGGTCTGAGGTTGAC	171.9
QY	1233	CCGAGCTTCTTGCCACCTTCCCTCC - GGCACTCAGCTCTGCATTCATCCCTCTTTAA	129.3
Db	1720	CCGAGCTTCTTGCCACCTTCCCTCCGCGGAGTCAAGCTCTCATCCATCCCTCTTTAA	177.9
QY	1292	TCTATGAATCTATAGGCTCGGTGTGTGTACGA - ACACACCCCTATCGTGTGCTTCAAT	135.0
Db	1780	TCTATGAATCTATAGGCTCGGTGTGTGTACGA - ACACACCCCTATCGTGTGCTTCAAT	183.9
QY	1351	ACTCAGCATTAACCATTTGGTTGAGGCGAAATTCAGAGCTTTCTCAATCAGATTATTAATC	141.0
Db	1840	ACTCAGCATTAACCATTTGGTTGAGGCGAAATTCAGAGCTTTCTCAATCAGATTATTAATC	189.9
QY	1411	TCCATTTTCATTAACGGGGAAACATCCCGCAGCCACTAGTGTGCTTTGTCACTGAA	147.0
Db	1900	TCCATTTTCATTAACGGGGAAACATCCCGCAGCCACTAGTGTGCTTTGTCACTGAA	195.9
QY	1471	GGTTAGATCTGAACCCAGGGGTGTCAACNGCTGCTCTCACTTCCCACTCTGGGGCATGA	153.0
Db	1960	GGTTAGATCTGAACCCAGGGGTGTCAACNGCTGCTCTCACTTCCCACTCTGGGGCATGA	201.9
QY	1531	GGAGATTTTCCCTCATATCTACCTCTCTAAAGGCTATGACCCCTCCACAGCTTCCAGC	159.9
Db	2020	GGAGATTTTCCCTCATATCTCTAAGCTCTCTAAAGGCTATGACCCCTCCACAGCTTCCAGC	207.9
QY	1591	TGGGGGATGGGGGAGTCAATAGGAAAGCCCCCATCTGCCATCTGGGATAGGAGACCTTCC	165.0
Db	2080	TGGGGGATGGGGGAGTCAATAGGAAAGCCCCCATCTGCCATCTGGGATAGGAGACCTTCC	213.9
QY	1651	ATCAGCCCTTAACCCCGGGGAAATGCGCTGTGCGCCAGAGCACTTTGGTTGTCCTCCAC	171.0
Db	2140	ATCAGCCCTTAACCCCGGGGAAATGCGCTGTGCGCCAGAGCACTTTGGTTGTCCTCCAC	219.9
QY	1711	ATACAGAAAGCAGGGTGAGAGGGGAAAGGATGGGTTCTCAGTTAGCAGAGGATCCCAAGGCAAG	177.0
Db	2200	ATACAGAAAGCAGGGTGAGAGGGGAAAGGATGGGTTCTCAGTTAGCAGAGGATCCCAAGGCAAG	225.9
QY	1771	TCAAGCTCTCCCTCCATAGCTCTCTGCTCAGTGTGAGTGTAGGAGTGGGCTCTCACTCCCA	183.0
Db	2260	TCAAGCTCTCTCCCTCCATAGCTCTCTGCTCAGTGTGAGTGTAGGAGTGGGCTCTCACTCCCA	231.9
QY	1831	CCACTCTTGGG - CCTTTGGGGGAGCACTTGGGAGGGGCGCTGTGAGAGACCTCTGACGCTGG	188.9
Db	2320	CCACTCTTGGGCGCTTTGGGGGAGCACTTGGGAGGGGCGCTGTGAGAGACCTCTGACGCTGG	237.9
QY	1890	AACCTGTATACACATTAAGCAGACTTCAC	192.0
Db	2380	AACCTGTATACACATTAAGCAGACTTCAC	241.0

```

US-09-918-995-34469
; Sequence 34469, Application US/09918995
; Publication No. US20030075623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34469
; LENGTH: 421
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-34469

```

Query Match	20.8%	Score 414.2	DB 10	Length 421
Best Local Similarity	99.3%	Pred. No. 6.5E-123		
Matches 416	Conservative 0	Mismatches 3	Indels 0	Gaps 0
QY	457	GGACATCTCAGCCAGGTTCTAGCCCCACGCACCCACAGACCCACAGGTGACACATCTTCC	516	
Db	3	GGACATCTCAGCCAGGTTCTAGCCCCACGCACCCACAGACCCACAGGTGACACATCTTCC	62	
QY	517	CAACTAGGGGCGCTCCACCTCATCCAGGGCAGGCCAGGGACTCCCTGGCTGACACATGA	576	
Db	63	CAACTAGGGGCGCTCCACCTCATCCAGGGCAGGCCAGGGACTCCCTGGCTGACACATGA	122	
QY	577	TGCCAGATTTCAGATTGTGGCTCCGTCACTTAAATCCAGACTACAGGGGCTGGGGTCAGG	636	
Db	123	TGCCAGATTTCAGATTGTGGCTCCGTCACTTAAATCCAGACTACAGGGGCTGGGGTCAGG	182	
QY	637	GAAAGAAATTTAAGAACCACCTCTGGGTGAGGGGAAATGGGACACAGAGACATATAGG	696	
Db	183	GAAAGAAATTTAAGAACCACCTCTGGGTGAGGGGAAATGGGACACAGAGACATATAGG	242	
QY	697	CAAGCTCTGCAGACAGACAGACAGACAAACCTCTGATCTATGATAGTCTGTGCAGGGCA	756	
Db	243	CAAGCTCTGCAGACAGACAGACAGACAAACCTCTGATCTATGATAGTCTGTGCAGGGCA	302	
QY	757	AGGGGACCAAGGACCTTGGAACTCTTTGGCCAAAGGGAGTGGAGACACAGAGGAAGT	816	
Db	303	AGGGGACCAAGGACCTTGGAACTCTTTGGCCAAAGGGAGTGGAGACACAGAGGAAGT	362	
QY	817	CACAGGCAAGGGTGCCTATCTTAATGTGAATATTTGCCGAGGGCTGACGAAGGCCAAG	875	
Db	363	CACAGGCAAGGGTGCCTATCTTAATGTGAATATTTGCCGAGGGCTGACGAAGGCCAAG	421	

RESULT 7

```

RESULT 8
US-09-918-995-5223
? Sequence 5223, Application US/09918995
? Publication No. US2003007623A1
? GENERAL INFORMATION:
? APPLICANT: Hyseq, Inc.
? TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
? TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
? FILE REFERENCE: 20411-756
? CURRENT APPLICATION NUMBER: US/09/918,995
? CURRENT FILING DATE: 2001-07-30
? PRIOR APPLICATION NUMBER: US/09/235,076
? PRIOR FILING DATE: 1999-01-20
? NUMBER OF SEQ ID NOS: 38054
? SOFTWARE: FASTSEQ for Windows Version 3.0
? SEQ ID NO 5223
? LENGTH: 412
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-918-995-5223

```


RESULT 11

US-09-873-319-23/C
 ; Sequence 23, Application US/09873319A
 ; Publication No. US20030134324A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Mungar, William E.
 ; APPLICANT: Kulkarni, Prakash
 ; APPLICANT: Getzenberg, Robert H.
 ; APPLICANT: Waga, Iwao
 ; APPLICANT: Yamamoto, Jun
 ; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic
 ; TITLE OF INVENTION: Hyperplasia Using Gene Expression Profiles
 ; FILE REFERENCE: 44921-5029-US
 ; CURRENT APPLICATION NUMBER: US/09/873,319A
 ; EARLIER APPLICATION NUMBER: US 60/223,323
 ; EARLIER FILING DATE: 2000-08-07
 ; NUMBER OF SEQ ID NOS: 755
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 23
 ; LENGTH: 402
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20030134324A1 AA045503
 US-09-873-319-23

Query Match 20.2%; Score 402; DB 10; Length 402;
 Best Local Similarity 100.0%; Pred. No. 5.6e-119; Indels 0; Gaps 0;
 Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1521 TGGGCACTGAGAGATTTCCCTCATTTCTACCTCTCTAGAGCTATGACCCCTCCAC 1580
 DB 402 TGGGCACTGAGAGATTTCCCTCATTTCTACCTCTCTAGAGCTATGACCCCTCCAC 343
 QY 1581 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 1640
 DB 342 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 283
 QY 1641 GGGACCTTCATCAGCTTACCTTCGGAATGCTGCTGCCCCAGTACTTGGTTT 1700
 DB 282 GGGACCTTCATCAGCTTACCTTCGGAATGCTGCTGCCCCAGTACTTGGTTT 223
 QY 1701 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 1760
 DB 222 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 163
 QY 1761 CGAGGCAAGTCAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1820
 DB 162 CGAGGCAAGTCAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 103
 QY 1821 CTCACTCCCACTCTGAGGCTTGGGGGAGACTGGGGAGGGGGCGGTGGAGAGCC 1880
 DB 102 CTCACTCCCACTCTGAGGCTTGGGGGAGACTGGGGAGGGGGCGGTGGAGAGCC 43
 QY 1881 TGAAGCTGGAAGCTGTATACAAATTAAGAGAGAGCTTACAG 1922
 DB 42 TGAAGCTGGAAGCTGTATACAAATTAAGAGAGAGCTTACAG 1

RESULT 12

US-09-873-367C-532/C
 ; Sequence 532, Application US/09873367C
 ; Publication No. US20030165839A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Young, Paul
 ; APPLICANT: Soppet, Paul
 ; APPLICANT: Endress, Gregory
 ; APPLICANT: Augustus, Meena
 ; APPLICANT: Ebner, Reinhard
 ; APPLICANT: Carter, Kenneth
 ; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
 ; TITLE OF INVENTION: Signature Gene Sets

FILE REFERENCE: 689290-64

; CURRENT APPLICATION NUMBER: US/09/873,367C
 ; PRIOR FILING DATE: 2003-04-29
 ; PRIOR APPLICATION NUMBER: U.S. 60/236,891
 ; PRIOR FILING DATE: 2000-09-29
 ; PRIOR APPLICATION NUMBER: U.S. 60/236,842
 ; PRIOR FILING DATE: 2000-09-29
 ; PRIOR APPLICATION NUMBER: U.S. 60/244,867
 ; PRIOR FILING DATE: 2000-11-01
 ; PRIOR APPLICATION NUMBER: U.S. 60/245,084
 ; PRIOR FILING DATE: 2000-11-01
 ; NUMBER OF SEQ ID NOS: 1067
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 532
 ; LENGTH: 402
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-873-367C-532

Query Match 20.2%; Score 402; DB 10; Length 402;
 Best Local Similarity 100.0%; Pred. No. 5.6e-119; Indels 0; Gaps 0;
 Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1521 TGGGCACTGAGAGATTTCCCTCATTTCTACCTCTCTAGAGCTATGACCCCTCCAC 1580
 DB 402 TGGGCACTGAGAGATTTCCCTCATTTCTACCTCTCTAGAGCTATGACCCCTCCAC 343
 QY 1581 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 1640
 DB 342 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 283
 QY 1641 GGGACCTTCATCAGCTTACCTTCGGAATGCTGCTGCCCCAGTACTTGGTTT 1700
 DB 282 GGGACCTTCATCAGCTTACCTTCGGAATGCTGCTGCCCCAGTACTTGGTTT 223
 QY 1701 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 1760
 DB 222 GTCTTCCAGCTGGGGGAGTGGGGGAGTATAGAAAAAGCCCTTCCCTCTGGGATA 163
 QY 1761 CGAGGCAAGTCAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1820
 DB 162 CGAGGCAAGTCAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 103
 QY 1821 CTCACTCCCACTCTGAGGCTTGGGGGAGACTGGGGAGGGGGCGGTGGAGAGCC 1880
 DB 102 CTCACTCCCACTCTGAGGCTTGGGGGAGACTGGGGAGGGGGCGGTGGAGAGCC 43
 QY 1881 TGAAGCTGGAAGCTGTATACAAATTAAGAGAGAGCTTACAG 1922
 DB 42 TGAAGCTGGAAGCTGTATACAAATTAAGAGAGAGCTTACAG 1

RESULT 13

US-09-918-995-3808
 ; Sequence 3808, Application US/0918995
 ; Publication No. US20030073623A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hyseq, Inc.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 ; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
 ; FILE REFERENCE: 20411-756
 ; CURRENT APPLICATION NUMBER: US/09/918,995
 ; PRIOR FILING DATE: 2001-07-30
 ; PRIOR APPLICATION NUMBER: US/09/235,076
 ; PRIOR FILING DATE: 1999-01-20
 ; NUMBER OF SEQ ID NOS: 38054
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 3808
 ; LENGTH: 486
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature

ORGANISM: Homo sapiens
 FEATURE:
 ; OTHER INFORMATION: MAP TO AI096775.9
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.7
 ; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.5
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.9
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.1
 ; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
 US-09-864-761-2033

Query Match 18.9%; Score 375; DB 9; Length 458;
 Best Local Similarity 94.0%; Freq. No. 3.2e-110;
 Matches 424; Conservative 0; Mismatches 20; Indels 7; Gaps 3;

OY	226	CACAAGTCCAGACTGCCGAGGAGCGTGAACCCGAGCGTGGTGAACCTGGCTACTGT	285
DB	444	CACAAGTCCAG-CTGCCGAGGAGCGTGAACCC--GACGTGGTGAACCTGG--CTCT	392
OY	286	CGGAGAGCGAGCAAGCTCTAGTCCGAGGAGCGGTACCCGATACAGAACAGTCCA	345
DB	391	GCGGAGAGCGAGCAAGCTCTAGTCCGAGGAGCGGTACCCGATACAGAACAGTCCA	332
OY	346	GCAGACCTTCGAGCGGAGATCAAAATCTAGTGGCTCTCCCTTCCCTTCCCTGTC	405
DB	331	GCAGACCTTCGAGCGGAGATCAAAATCTAGTGGCTCTCCCTTCCCTTCCCTGTC	272
OY	406	CGGCCCCAGCGCTGCGAGCAAGCTGCTAAACCCATTACACAGCTCCAGACATCTC	465
DB	271	CGGCCCCAGCGCTGCGAGCAAGCTGCTAAACCCATTACACAGCTCCAGACATCTC	212
OY	466	AGCCAGGTTCTAGCCCCCAGGAGCCAGACCCAGGAGGAGGAGGAGGAGGAGGAGG	525
DB	211	AGCCAGGTTCTAGCCCCCAGGAGCCAGACCCAGGAGGAGGAGGAGGAGGAGGAGG	152
OY	526	GCCCTCCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	585
DB	151	GCCCTCCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	92
OY	586	TTCAAGTTTGGCTCTCCGTCATTATCCAGAGTACAGGGGCTGGGGTCAAGGAAGAGA	645
DB	91	TTCAAGTTTGGCTCTCCGTCATTATCCAGAGTACAGGGGCTGGGGTCAAGGAAGAGA	32
OY	646	TCTAAGAACCCACTGTGGGTCAAGGGGAATG	676
DB	31	TCTAAGAACCCACTGTGGGTCAAGGGGAATG	1

Search completed: October 24, 2004, 16:00:21
 Job time : 974 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: October 24, 2004, 11:30:51 ; Search time 77 Seconds

(without alignments)
708.141 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848
Sequence: 1 MPRLYIDGVFPIQLVREKLH.....LSQIRPTISIFINGETSPSH 152

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : AAgeneSeq_23sep04:*
1: geneseqp1980s:*
2: geneseqp1980s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	846	99.8	152	6	ABO23236 Human bre
2	846	99.8	152	8	ADP66735 Novel hum
3	846	99.8	152	8	ADO60245 Human NOV
4	836.5	98.6	151	4	AAB61132 Human NOV
5	836.5	98.6	151	7	ADM56371 Human cel
6	739	87.1	136	8	AD119772 Human NOV
7	500.5	59.0	102	2	AAV48502 Human NOV
8	312	36.8	101	6	AAB61152 Protein e
9	312	36.8	101	6	ABO23256 Human bre
10	312	36.8	101	8	ADP66803 Novel hum
11	312	36.8	101	8	ADP66803 Novel hum
12	305	36.0	101	7	ADM56440 Human cel
13	296	34.9	101	8	AD119840 Human NOV
14	113.5	13.4	703	5	ABR05728 Human tita
15	95	11.2	247	7	ADB64988 Human pro
16	90.5	10.7	502	8	ABO59296 Human gen
17	90.5	10.7	665	8	ADN99681 Novel hum
18	88	10.4	200	5	ADK34802 Novel hum
19	88	10.4	1745	4	AAQ77793 Human pro
20	88	10.4	1745	5	ABR97234 Novel hum
21	88	10.4	1745	8	ADQ19841 Human sof
22	85	10.0	181	8	ADG22767 Cyanophag
23	84.5	10.0	232	5	ABP41247 Human ova
24	84.5	10.0	334	7	ABO67883 Pseudomon
25	83.5	9.8	631	5	ABG32473 Human pro

26	83	9.8	147	4	AAU32381 Novel hum
27	82.5	9.7	68	4	AAQ74508 Human col
28	82.5	9.7	884	2	AAW59666 Amino aci
29	82.5	9.7	884	2	AAW94074 Human G-P
30	81.5	9.6	457	7	ABO80925 Pseudomon
31	81	9.6	1739	4	AAQ77792 Murine pr
32	80.5	9.5	1554	4	ABR11890 Human pro
33	80.5	9.5	3298	4	AAE03657 Human ext
34	80.5	9.5	3312	7	ADJ69693 Human hea
35	80.5	9.5	3312	8	ADO29247 Human GPC
36	80.5	9.5	3312	8	ADO17551 Human sof
37	80.5	9.5	3317	8	ADH18895 Human cel
38	80.5	9.5	4115	5	ABP53570 Human NOV
39	80	9.4	361	7	ADP94865 Human gen
40	80	9.4	365	8	ADL71718 Novel hum
41	80	9.4	542	7	ADP94893 Human gen
42	80	9.4	542	8	ADL71719 Novel hum
43	80	9.4	583	7	ADJ70094 Human hea
44	80	9.4	920	6	ABP70827 Human C1Q
45	80	9.4	1016	3	ABR41524 Human ORF

ALIGNMENTS

RESULT 1
ID ABO23236
ID ABO23236 standard, protein, 152 AA.
XX
AC ABO23236;
XX
DT 03-SEP-2003 (first entry)
XX
DE Human breast tumour associated protein 47-like polypeptide NOV4.
XX
XX Tissue typing; cancer; breast cancer; colon cancer; lung cancer; sarcoma;
XX pancreatic cancer; uterine cancer; organ transplantation disorder;
XX cardiovascular disease; melanoma; atherosclerosis; diabetes mellitus;
XX ischaemic heart disease; haemorrhage; peripheral vascular disease;
XX thrombosis; hypertension; systemic lupus erythematosus; haematopoiesis;
XX tissue regeneration; wound healing; hyperproliferative disorder;
XX psoriasis; neural disorder; Parkinson's disease; Alzheimer's disease;
XX Huntington's disease; multiple sclerosis; amyotrophic lateral sclerosis;
XX ALS; peripheral neuropathy; nervous system tumour; neurotoxin; tremor;
XX neuropathy; acute brain injury; peripheral nerve trauma; human; NOVX;
XX gene therapy; epilepsy; breast tumour associated protein 47.
XX
XX Homo sapiens.
XX
XX US2003027158-A1.
XX
PD 06-FEB-2003.
XX
XX 15-OCT-2001; 2001US-00977418.
XX
XX 03-JUN-1999; 99US-0137322P.
XX
XX 16-MAR-2000; 2000US-0189810P.
XX
XX 22-MAR-2000; 2000US-0191158P.
XX
XX 30-MAR-2000; 2000US-0193086P.
XX
XX 03-MAY-2000; 2000US-0201388P.
XX
XX 31-MAY-2000; 2000US-00584411.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinketsu RA, Fernandes E, Herrman J, Vernet C;
XX WPI: 2003-492028/46.
XX
XX N-PSDB; ACDB0257.
XX
XX New nucleic acid sequence encoding a human breast tumor-associated
XX protein 47-like polypeptide, useful for treating cardiovascular
XX disorders, neural disorders, diabetes mellitus and cancers.
XX

PS Claim 1; Page 16-17; 100pp; English.

XX The invention relates to a new isolated NOV4 nucleic acid. The nucleic acid is useful for identifying a compound that binds the nucleic acid. The nucleic acid is useful in gene therapy, in screening assays, in detection assays e.g. chromosomal mapping, cell and tissue typing and forensic biology, predictive medicine e.g. diagnostic assays, prognostic assays, monitoring clinical trials and pharmacogenomics and methods of treatment including therapeutic and prophylactic. The nucleic acid is also useful for expressing NOVX protein. The nucleic acid is also useful to provide polynucleotide reagents e.g. labelled probes that are useful in an in situ hybridisation technique, for identifying a specific tissue (for example brain tissue) and for use in forensic science. The nucleic acid is also useful for mapping genes on a chromosome and thus locating gene regions associated with genetic disease, identifying an individual from a minute biological sample and to aid in forensic identification of biological sample. The nucleic acid is also useful for treating cancer, especially cancers of the breast, colon, lung, pancreas or uterus, or a melanoma or sarcoma. The nucleic acid is also useful for treating disorders related to organ transplantation, cardiovascular diseases, atherosclerosis, ischaemic heart disease, haemorrhage, diabetes mellitus, peripheral vascular disease, thrombosis, hypertension and systemic lupus erythematosus. NOVX protein encoded by the nucleic acid is useful for regulating haematopoiesis, for regeneration of bone, cartilage, tendon ligament and/or nerve tissue growth or regeneration and for wound healing. The nucleic acid is also useful for treating infections, hyperproliferative disorders e.g. psoriasis, and neural disorders, including Parkinson's disease, Alzheimer's disease, Huntington's disease, multiple sclerosis, amyotrophic lateral sclerosis (ALS), peripheral neuropathy, tumours of the nervous system, exposure to neurotoxins, acute brain injury, peripheral nerve trauma or injury and other neuropathies, epilepsy and/or tremors. The present sequence represents the amino acid sequence of a human breast tumour associated protein 47-like polypeptide

XX Sequence 152 AA;

Query Match 99.8%; Score 846; DB 6; Length 152;

Best Local Similarity 100.0%; Pred. No. 3,3e-84; Mismatches 0; Indels 0; Gaps 0;

Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPELYIDGVFPPIQVLEKHEPVALRLIGPGKAKMGKMGMSFSLPQACVXGAAPMOT 60
DB 1 MPELYIDGVFPPIQVLEKHEPVALRLIGPGKAKMGKMGMSFSLPQACVXGAAPMOT 60
QY 61 GSNRDSGQMDALPPRGWTPASCHLPISALHPSPSLIESIGSVCTTHPRCSN 120
DB 61 GSNRDSGQMDALPPRGWTPASCHLPISALHPSPSLIESIGSVCTTHPRCSN 120
QY 121 TQHYHMLRPNSELSQIRFTISIFINGETSPSH 152
DB 121 TQHYHMLRPNSELSQIRFTISIFINGETSPSH 152

RESULT 2

AD66735 ADF66735 standard; protein; 152 AA.

AC ADF66735;

DT 12-FEB-2004 (first entry)

XX Novel human protein NOV4.

XX cytosolic; hepatotropic; vulnery; antipsoriatic; osteopathic;
XX antitubercular; antiatherosclerotic; haemostatic; vasoregic;
XX thrombolytic; antidiabetic; hypotensive; dermatological;
XX immunosuppressive; antineoplastic; immunostimulant; fungicide;
XX virucide; proteoasidic; neuroprotective; antineuritic; antitubercular;
XX antistatic; antiparkinsonian; nootropic; anticonvulsant;
XX NOVX modulator; cancer; hyperproliferative disease; cirrhosis; keloid;
XX psoriasis; tissue hypertrophy; osteoarthritis;
XX atherosclerotic plaque formation; haemorrhage; ischaemic disease;
XX thrombosis; diabetes mellitus; hypertension; hypothyroidism;

KW immune deficiency; severe combined immunodeficiency; SCID; infection;
KW malaria; candidiasis; autoimmune disorder; connective tissue disease;
KW multiple sclerosis; systemic lupus, erythematosus; rheumatoid arthritis;
KW autoimmune pulmonary inflammation; Guillain-Barre syndrome;
KW autoimmune thyroiditis; insulin dependent diabetes mellitus;
KW myasthenia gravis; graft- versus-host disease;
KW autoimmune inflammatory eye disease; asthma; haematopoiesis;
KW tissue regeneration; wound healing; tissue repair; burn; incision; ulcer;
KW periodontal disease; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
KW human.

OS Homo sapiens.

PN US2003199103-A1.

PD 23-OCT-2003.

PF 15-OCT-2001; 2001US-00977639.

PR 03-JUN-1999; 99US-0137322P.

PR 16-MAR-2000; 2000US-0189810P.

PR 22-MAR-2000; 2000US-0191158P.

PR 30-MAR-2000; 2000US-0193086P.

PR 03-MAY-2000; 2000US-0201388P.

PR 31-MAY-2000; 2000US-00584411.

PA (CURA-) CURAGEN CORP.

PI Shinkets RA, Fernandes E, Heitman J, Vernet C;

PT WPI; 2004-021196/02.

PR N-PSDB; ADF66734.

PS Disclosure; SEQ ID NO 8; 165pp; English.

XX The invention describes a substantially purified polypeptide (I) having amino acid sequence chosen from a fully defined NOVX sequence (SI) of 708 amino acids as given in the specification, or polypeptide having one or more conservative amino acid substitutions of (SI) or mutant or variant of (SI). (I) having (SI) is useful for diagnosing a pathological condition associated with (I) or its activity in a subject e.g. cancer. (I) useful in treatment of cancer, hyperproliferative diseases, cirrhosis, keloid, psoriasis, tissue hypertrophy, osteoarthritis, atherosclerotic plaque formation, haemorrhage, ischaemic heart or renal disease, thrombosis, diabetes mellitus, hypertension, hypothyroidism. (I) is useful in treatment of various immune deficiencies and disorders such as severe combined immunodeficiency (SCID), bacterial infection, viral infection such as herpes viral infection, protozoan infection such as malaria, fungal infection such as candidiasis. (I) is also useful in treating autoimmune disorders such as connective tissue disease, multiple sclerosis, systemic lupus, erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft- versus-host disease and autoimmune inflammatory eye disease and asthma. (I) useful in regulation of haematopoiesis, regeneration and tissue growth of bone, cartilage, tendon, ligament and useful for wound healing and tissue repair. (I) is also useful in treatment of burns, incisions and ulcers. (I) also useful in treatment of periodontal disease, Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. (I) has effective antitumour and antiinflammatory activity. This is the amino acid sequence of a novel human NOVX protein.

XX Sequence 152 AA;

Query Match 99.8%; Score 846; DB 8; Length 152;

Best Local Similarity 100.0%; Pred. No. 3,3e-84; Mismatches 0; Indels 0; Gaps 0;

Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPHLYIDGVFPPIQLVREKLEHETPAVRLILGPGKAKWGMGWSFSLPQACVXGAAPMOT 60
DB 1 MPHLYIDGVFPPIQLVREKLEHETPAVRLILGPGKAKWGMGWSFSLPQACVXGAAPMOT 60
QY 61 GSPNDSGSGMDGALGPRGWTPTASCHLPLRQSLHPSPSLIYESIGVCVTHPRCPSPN 120
DB 61 GSPNDSGSGMDGALGPRGWTPTASCHLPLRQSLHPSPSLIYESIGVCVTHPRCPSPN 120
QY 121 TQHYHMLRPNSELSQLRFTISIFINGETSPSH 152
DB 121 TQHYHMLRPNSELSQLRFTISIFINGETSPSH 152

RESULT 3
ADO60245 standard; protein; 152 AA.
ID ADO60245; protein; 152 AA.
AC ADO60245;
XX
DT 15-JUL-2004 (first entry)
XX
DE Human NOVA4 protein.
XX
KM Human; NOVA4 protein; cancer; hyperproliferative disease; cirrhosis;
KM keloid; psoriasis; tissue hypertrophy; osteoarthritis;
KM atherosclerotic plaque formation; haemorrhage; ischaemic heart disease;
KM renal disease; thrombosis; diabetes mellitus; hypertension;
KM hypothyroidism; severe combined immunodeficiency; SCID; infection;
KM malaria; candidiasis; autoimmune disorder; connective tissue disease;
KM multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KM autoimmune pulmonary inflammation; Guillain-Barre syndrome;
KM autoimmune thyroiditis; myasthenia gravis; graft-versus-host disease;
KM autoimmune inflammatory eye disease; asthma; burn; incision; ulcer;
KM periodontal disease; Alzheimer's disease; Parkinson's disease;
KM Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
KM haematopoiesis; wound healing; tissue repair; antitumor;
KM antinflammatory.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..25
FT /label= Signal_peptide
FT Protein 26..152
FT /note= "Mature NOVA4 protein"
FT Misc-difference 53
FT /note= "Encoded by CNT"
XX
US2003134430-A1.
XX
PD 17-JUL-2003.
XX
PF 15-OCT-2001; 2001US-0097751.
XX
PR "03-JUN-1999; 99US-0137322P.
PR 16-MAR-2000; 2000US-0189810P.
PR 22-MAR-2000; 2000US-0191158P.
PR 30-MAR-2000; 2000US-0193086P.
PR 03-MAY-2000; 2000US-0201388P.
PR 31-MAY-2000; 2000US-00584411.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Fernandes E, Herriman J, Vernet C;
XX WPI; 2004-068928/07.
XX N-PSDB; ADO60244.
XX
PT Novel substantially purified NOVA4 polypeptide for treating severe
PT combined immunodeficiency, candidiasis, cancer, asthma, multiple
XX sclerosis, systemic lupus erythematosus.
XX

PS Disclosure; SEQ ID NO 8; 155pp; English.
XX
CC The invention relates to human NOVA4 polypeptides and polynucleotides.
CC NOVA4 sequences are useful in the treatment of cancer, hyperproliferative
CC diseases, cirrhosis, keloid, psoriasis, tissue hypertrophy,
CC osteoarthritis, atherosclerotic plaque formation, haemorrhage, ischaemic
CC heart or renal disease, thrombosis, diabetes mellitus, hypertension,
CC hypothyroidism, asthma, burns, incisions, ulcers, periodontal disease,
CC Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, Shy-Drager syndrome, immune deficiencies
CC and disorders such as severe combined immunodeficiency (SCID), bacterial
CC infection, viral infection e.g. herpes viral infection, protozoan
CC infection e.g. malaria, fungal infection e.g. candidiasis, autoimmune
CC disorders such as connective tissue disease, multiple sclerosis, systemic
CC lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
CC inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin
CC dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease
CC and autoimmune inflammatory eye disease. The invention is useful in
CC regeneration and tissue growth of bone, cartilage, tendon, ligament,
CC haematopoiesis regulation, wound healing and tissue repair. Sequences of
CC the invention also exhibit antitumor and antinflammatory activities. The
XX present sequence is human NOVA4 protein.
XX
SQ Sequence 152 AA;
XX
Query Match 99.8%; Score 846; DB 8; Length 152;
Best Local Similarity 100.0%; Pred. No. 3.3e-84;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPHLYIDGVFPPIQLVREKLEHETPAVRLILGPGKAKWGMGWSFSLPQACVXGAAPMOT 60
DB 1 MPHLYIDGVFPPIQLVREKLEHETPAVRLILGPGKAKWGMGWSFSLPQACVXGAAPMOT 60
QY 61 GSPNDSGSGMDGALGPRGWTPTASCHLPLRQSLHPSPSLIYESIGVCVTHPRCPSPN 120
DB 61 GSPNDSGSGMDGALGPRGWTPTASCHLPLRQSLHPSPSLIYESIGVCVTHPRCPSPN 120
QY 121 TQHYHMLRPNSELSQLRFTISIFINGETSPSH 152
DB 121 TQHYHMLRPNSELSQLRFTISIFINGETSPSH 152

RESULT 4
AAB61132 standard; protein; 151 AA.
ID AAB61132 standard; protein; 151 AA.
AC AAB61132;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human NOVA4 protein.
XX
KM Human; NOVA4 protein; cancer; hyperproliferative disease; cirrhosis;
KM keloid; psoriasis; tissue hypertrophy; osteoarthritis;
KM atherosclerotic plaque formation; haemorrhage; ischaemic heart or renal disease; thrombosis; diabetes mellitus; hypertension;
KM hypothyroidism; severe combined immunodeficiency; SCID; infection;
KM malaria; candidiasis; autoimmune disorder; connective tissue disease;
KM multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KM autoimmune pulmonary inflammation; Guillain-Barre syndrome;
KM autoimmune thyroiditis; myasthenia gravis; graft-versus-host disease;
KM autoimmune inflammatory eye disease; asthma; burn; incision; ulcer;
KM periodontal disease; Alzheimer's disease; Parkinson's disease;
KM Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
KM haematopoiesis; wound healing; tissue repair; antitumor;
KM antinflammatory.
XX
OS Homo sapiens.
XX
PN WO200075321-A2.
XX
PD 14-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US015303.
XX
PR "03-JUN-1999; 99US-0137322P.
PR 16-MAR-2000; 2000US-0189810P.
PR 22-MAR-2000; 2000US-0191158P.
PR 30-MAR-2000; 2000US-0193086P.
PR 31-MAY-2000; 2000US-00137322.
XX
PA (CURA-) CURAGEN CORP.
XX

PI Shinkets RA, Fernandes E, Herrman J, Vernet C;
XX
XX WPI: 2001-102403/11.
DR N-PSDB; AAF27852.
XX
PT New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
PT diagnostic marker, protein therapeutic, antibody or small molecule drug
PT target for treating immune, proliferative and metabolic diseases and
PT wound healing.
XX
PS Claim 1; Page 26-27; 194pp; English.
XX
XX The present sequence is a new isolated polypeptide (NOVX). The NOVX
CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
CC treating or preventing NOVX-associated disorders. They are also useful
CC for determining the presence of or a predisposition to a disease
CC associated with altered levels of the NOVX polypeptide or nucleic acid.
CC These NOVX-associated disorders include hyperplasia, tumours,
CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
CC are especially useful in gene therapy. Specifically, NOVX is useful as a
CC diagnostic marker or prognostic marker, protein therapeutic and antibody
CC target or small molecule drug target to treat disorders in the immune
CC response pathway, thyroid and metabolic diseases, bone metabolic
CC disorders, diseases of the pancreas (e.g. diabetes or digestive
CC disorders), proliferative diseases, or tissue regeneration and
CC development (e.g. wound healing or treatment of burns)
XX
SQ Sequence 151 AA;
SQ
Query Match 98.6%; Score 836.5; DB 4; Length 151;
Best Local Similarity 99.3%; Pred. No. 3,6e-83;
Matches 151; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 MPELYIDGVFPPIQVLEKLEHETPAVLRLLGPPGKAKWGMGWSFSLPQACVGAAPMOT 60
Db 1 MPELYIDGVFPPIQVLEKLEHETPAVLRLLGPPGKAKWGMGWSFSLPQACV-GAAPMOT 59
QY 61 GSPNRDGSQMDGALGPRGWTASCHLPIRQALHPSLSLYESIGSVCTTHPRCPSN 120
Db 60 GSPNRDGSQMDGALGPRGWTASCHLPIRQALHPSLSLYESIGSVCTTHPRCPSN 119
QY 121 TCHYHWRNPSELSQIRFTISIFINGETSPSH 152
Db 120 TCHYHWRNPSELSQIRFTISIFINGETSPSH 151
RESULT 5
ADMS6371
ID ADM56371 standard; protein; 151 AA.
XX
AC ADM56371;
XX
XX 03-JUN-2004 (first entry)
XX
XX Human cell adhesion molecule NOV4 #1.
XX
XX Human: cell adhesion molecule; NOVX; Cancer; leukaemia; lymphoma;
XX melanoma; neurological disorder; epilepsy; Alzheimer's disease;
XX ischaemic cerebrovascular disease; stroke; Alzheimer's disease;
XX Pick's disease; vesicular transport disorder; cystic fibrosis;
XX diabetes mellitus; Grave's disease; goiter; gastrointestinal disorder;
XX ulcerative colitis; gastric ulcer; duodenal disorder; autoimmune disease;
XX allergic reaction; autoimmune haemolytic anemia; rheumatoid arthritis;
XX viral infection; bacterial infection; fungal infection;
XX helminthic infection; protozoal infections.
XX
XX Homo sapiens.
XX
XX US2003082554-A1.
XX
XX 01-MAY-2003.

XX
XX 15-OCT-2001; 2001US-00977033.
XX
XX 03-JUN-1999; 99US-0137322P.
XX
XX 16-MAR-2000; 2000US-0189810P.
XX
XX 22-MAR-2000; 2000US-0191159P.
XX
XX 30-MAR-2000; 2000US-0193086P.
XX
XX 03-MAY-2000; 2000US-0201388P.
XX
XX 31-MAY-2000; 2000US-00584411.
XX
XX (CUBA-) CUBAGEN CORP.
XX
XX Shinkets RA, Fernandes E, Herrman J, Vernet C;
XX
XX WPI: 2003-616079/58.
XX
XX N-PSDB; ADM56372.
XX
XX New nucleic acids encoding human cell adhesion molecule-like proteins,
XX useful for treating e.g. cancers, neurological disorders, viral,
XX bacterial, fungal, helminthic and protozoal infections.
XX
XX Example 3; SEQ ID NO 8; 78pp; English.
XX
XX The invention relates to an isolated nucleic acid encoding a human cell
XX adhesion molecule-like protein, comprising a sequence encoding a
XX polypeptide having a sequence appearing as ADM56387, a sequence at least
XX 90% identical to the nucleic acid, a sequence encoding a polypeptide
XX having conservative amino acid substitutions to the protein or a fragment
XX comprising at least 20 nucleotides. Also included are an oligonucleotide
XX sequence that is complementary to (and hybridises under stringent
XX conditions with) the nucleic acid (or a portion of it), a vector
XX comprising the nucleic acid, a cell comprising the vector, a
XX pharmaceutical composition comprising the nucleic acid and a
XX process for identifying a compound that binds the nucleic acid, and a
XX compound identified by the process. Disclosed as new are the cDNA and
XX proteins for novel cell adhesion molecules (termed NOVX, being NOV1-23).
XX
XX The NOVX polypeptide, nucleic acid or antibody are useful in the
XX manufacture of a medicament for treating a syndrome associated with a
XX human disease selected from NOVX-associated disorder, such as cancers
XX (e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle,
XX ovary, testis and uterus), neurological disorders (e.g. epilepsy,
XX ischaemic cerebrovascular disease, stroke, Alzheimer's disease or Pick's
XX disease), disorders of vesicular transport (e.g. cystic fibrosis,
XX diabetes mellitus, Grave's disease, or goiter), gastrointestinal
XX disorders (e.g. ulcerative colitis, or gastric and duodenal disorders),
XX autoimmune diseases (e.g. allergic reactions, autoimmune haemolytic
XX anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic
XX and protozoal infections. The polypeptides can be used as immunogens to
XX produce antibodies and as vaccines. The sequences may further be used in
XX chromosome mapping, identifying individual from minute biological samples
XX (tissue typing), and in forensic identification of a biological sample.
XX
XX NOTE: The authors have mis-labelled the sequences as they appear on pages
XX 12-28 of the patent, it is clear from table 3, the examples and the
XX claims that the SEQ ID numbers for the cDNAs should be the odd numbers
XX from 1-45 and the proteins should be the even numbers from 2-46. The
XX present sequence represents a NOVX cell adhesion molecule of the
XX invention.
XX
XX Sequence 151 AA;
XX
SQ
Query Match 98.6%; Score 836.5; DB 7; Length 151;
Best Local Similarity 99.3%; Pred. No. 3,6e-83;
Matches 151; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 MPELYIDGVFPPIQVLEKLEHETPAVLRLLGPPGKAKWGMGWSFSLPQACVGAAPMOT 60
Db 1 MPELYIDGVFPPIQVLEKLEHETPAVLRLLGPPGKAKWGMGWSFSLPQACV-GAAPMOT 59
QY 61 GSPNRDGSQMDGALGPRGWTASCHLPIRQALHPSLSLYESIGSVCTTHPRCPSN 120
Db 60 GSPNRDGSQMDGALGPRGWTASCHLPIRQALHPSLSLYESIGSVCTTHPRCPSN 119

QY 121 TOHHMLRPNSLSQIRFTISIFINGETSPSH 152
 DB 120 TOHHMLRPNSLSQIRFTISIFINGETSPSH 151

RESULT 6
 AD119772
 ID AD119772 standard; protein: 136 AA.

XX AC AD119772;
 XX DT 22-APR-2004 (first entry)
 XX DE Human NOV4 protein.

XX Secreted protein; NOVX; diagnosis; metabolic disorder; diabetes; obesity;
 XX infection; anorexia; cancer; cardiovascular disease; hypertension;
 XX atherosclerosis; neurodegenerative disorder; Alzheimer's disease;
 XX Parkinson's disease; epilepsy; immune disorder; osteoarthritis;
 XX haematopoietic disorder; inflammatory skin disorder; asthma;
 XX dyslipidemia; neurogenesis; cell differentiation; cell proliferation;
 XX haematopoiesis; wound healing; angiogenesis; chromosome mapping;
 XX tissue typing; preventive medicine; pharmacogenomic; gene therapy;
 XX anorectic; cardiac; virucide; antibacterial; fungicide; protozoacide;
 XX neurotropic; neuroprotective; dermatologically; human.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 FT Peptide 1..54
 FT MISC-difference 52..53
 FT /note= "Encoded by GTTCNTGGG"
 FT Protein 55..136
 FT /note= "Human mature NOV protein"
 FT MISC-difference 60..61
 FT /note= "Encoded by bases from 1171 to 1221"
 FT
 XX US2004002134-A1.

XX PD 01-JAN-2004.
 XX PF 15-OCT-2001; 2001US-00977819.

XX PR 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.

XX PA (CUPA-) CURAGEN CORP.

XX PI Shinkets RA, Fernandes ER, Herrman JL, Vernet CAM;

XX DR WPI; 2004-070737/07.

XX DR N-PsDB; AD119771.

XX PT New NOVX nucleic acids encoding human K1A0768 protein-like and human
 XX PT protein PRO-228 polypeptides, useful for treating NOVX-associated
 XX PT disorders.

XX PS Disclousure; SEQ ID NO 8; 95pp; English.

XX The present invention is based in part on the discovery of novel secreted
 XX and membrane-bound polypeptides and their encoding polynucleotides. The
 XX nucleic acids and polypeptides are collectively referred as NOVX. The
 XX invention is useful for treating, preventing and diagnosing diseases such
 XX as metabolic disorders, diabetes, obesity, infectious diseases such as
 XX viral, bacterial, fungal, helminthic and protozoal infections, anorexia,
 XX cancer, cardiovascular diseases such as hypertension and atherosclerosis,
 XX neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,
 XX epilepsy, immune disorders such as osteoarthritis, haematopoietic
 XX disorders, inflammatory skin disorders, asthma and various dyslipidemias.

CC The invention is also useful as targets for the identification of small
 CC molecules that modulate or inhibit e.g. neurogenesis, cell
 CC differentiation, cell proliferation, haematopoiesis, wound healing and
 CC angiogenesis, as hybridisation probes, in chromosome mapping, tissue
 CC typing, preventive medicine and pharmacogenomics. The invention is also
 CC useful in gene therapy. The present sequence is human NOV protein.
 XX
 XX SQ Sequence 136 AA;

Query Match 87.1%; Score 739; DB 8; Length 136;
 Best Local Similarity 89.5%; Pred. No. 1.5e-72;
 Matches 136; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MPELYIDGVFPQLVREKLEHPAVLRLLGPPGKAKMGWGWMSFSLPQACVGAAPNQT 60
 DB 1 MPELYIDGVFPQLVREKLEHPAVLRLLGPPGKAKMGWGWMSFSLPQACV-GAAPNQT 59
 QY 61 GSPNRDGSQMDGALGPRGWTPTASCALPLRQSAALHPSLSLYESIGSVCTTHPIRCPSN 120
 DB 60 -----GPRGWTPTASCALPLRQSAALHPSLSLYESIGSVCTTHPIRCPSN 104
 QY 121 TOHHMLRPNSLSQIRFTISIFINGETSPSH 152
 DB 105 TOHHMLRPNSLSQIRFTISIFINGETSPSH 136

RESULT 7
 AA48502
 ID AA48502 standard; protein: 102 AA.

XX AC AA48502;

XX DT 08-DEC-1999 (first entry)

XX DE Human breast tumour-associated protein 47.

XX KW Expressed sequence tag; EST; human; breast; cancer; cytostatic;
 XX KW medicaments; gene therapy; treatment; fat metabolism.

XX OS Homo sapiens.

XX PN DE19813835-A1.

XX PD 23-SEP-1999.

XX PF 20-MAR-1998; 98DE-01013835.

XX PR 20-MAR-1998; 98DE-01013835.

XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenchal A;

XX DR WPI; 1999-528979/45.

XX DR N-PsDB; AA233583.

XX PT Human nucleic acid sequences and protein products from normal breast
 XX PT tissue, useful for breast cancer therapy.

XX PS Claim 28; 179; 206pp; German.

XX This invention describes novel human nucleic acid sequences from normal
 XX breast tissue which have cytosstatic activity. The nucleic acid sequences
 XX can be used to produce and isolate full-length gene sequences. They can
 XX be used to express proteins, which can be used as tools to find an
 XX activity against breast cancer. The sequences can be used in sense or
 XX antisense form. They are especially useful for medicaments for gene
 XX therapy to treat breast cancer and for treating illnesses associated with
 XX fat metabolism. AA48456-Y48539 represent protein fragments encoded by
 XX the expressed sequence tags described in the method of the invention

XX SQ Sequence 102 AA;

Query Match 59.0%; Score 500.5; DB 2; Length 102;
 Best Local Similarity 90.0%; Pred. No. 1,3e-46;
 Matches 90; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

QY 1 MPHLYIDGVFPPIQVLRKHEHPAVLRLLGPPGKAKMGWGMSFSLPQACVYGAAPMOT 60
 DB 1 MPHLYIDGVFPPIQVLRKHEHPAVLRLLGPPGKAKMGWGMSFSLPQACVYGAAPMOT 60

QY 61 GSPNRDSQSGMDGALGPRGWTTPASCHLPLRQSLHSPSLI 100
 DB 61 GSPNRDSQSGMDGALGPRGWTTPASCHLPLRAVS---SPSL 97

RESULT 8
 AAB61152
 ID AAB61152 standard; protein; 101 AA.
 AC AAB61152;
 DT 30-MAR-2001 (first entry)
 DE Protein encoded by human NOV4 cDNA clone 3189601.
 KW Human; NOVX; antiinflammatory; cytostatic; neuroprotective;
 KW cerebroprotective; immunomodulator; vulnerary; vasotropic; gene therapy;
 KW hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;
 KW diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.
 OS Homo sapiens.
 PN MO200075321-A2.
 XX 14-DEC-2000.
 PD 01-JUN-2000; 2000MO-US015303.
 PF 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 31-MAY-2000; 2000US-00137322.
 XX (CURA-) CURAGEN CORP.
 PA Shimkete RA, Fernandes E, Herrman J, Vernet C;
 PI WPI; 2001-102403/11.
 DR WPI; 2001-102403/11.
 XX New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
 PT diagnostic marker, protein therapeutic, antibody or small molecule drug
 PT target for treating immune, proliferative and metabolic diseases and
 PT wound healing.
 XX Example 3; Page 174; 194pp; English.
 PS The present sequence is a new isolated polypeptide (NOVX). The NOVX
 CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
 CC treating or preventing NOVX-associated disorders. They are also useful
 CC for determining the presence of or a predisposition to a disease
 CC associated with altered levels of the NOVX polypeptide or nucleic acid.
 CC These NOVX-associated disorders include hyperplasias, tumours,
 CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
 CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
 CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
 CC are especially useful in gene therapy. Specifically, NOVX is useful as a
 CC diagnostic marker or prognostic marker, protein therapeutic and antibody
 CC target or small molecule drug target to treat disorders in the immune
 CC response pathway, thyroid and metabolic diseases, bone metabolic
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive
 CC disorders), proliferative diseases, or tissue regeneration and
 CC development (e.g. wound healing or treatment of burns)
 XX Sequence 101 AA;

Query Match 36.8%; Score 312; DB 4; Length 101;
 Best Local Similarity 60.0%; Pred. No. 5.4e-26;
 Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

QY 1 MPHLYIDGVFPPIQVLRKHEHPAVLRLLGPPGKAKMGWGMSFSLPQACVYGAAPMOT 60
 DB 1 MPHLYIDGVFPPIQVLRKHEHPAVLRLLGPPGKATMGWGMSFSLPRVPLGLPCRQ 60

QY 61 GSPNRDSQSGMDG---ALGPRGWTTPASCHLPLRQSLHSPSLI 101
 DB 61 DHLTMEARAWMGLWLVLEVGQ---LLATFPQSQSLHSPSLI 101

RESULT 9
 ABO23256
 ID ABO23256 standard; protein; 101 AA.
 AC ABO23256;
 DT 03-SEP-2003 (first entry)
 DE Human breast tumour associated protein 47-like polypeptide NOV4 product.
 KW Tissue typing; cancer; breast cancer; colon cancer; lung cancer; sarcoma;
 KW pancreatic cancer; uterine cancer; organ transplantation disorder;
 KW cardiovascular disease; melanoma; atherosclerosis; diabetes mellitus;
 KW ischaemic heart disease; haemorrhage; peripheral vascular disease;
 KW thrombosis; hypertension; systemic lupus erythematosus; haematoplasias;
 KW tissue regeneration; wound healing; hyperproliferative disorder;
 KW psoriasis; neural disorder; Parkinson's disease; Alzheimer's disease;
 KW Huntington's disease; multiple sclerosis; amyotrophic lateral sclerosis;
 KW ALS; peripheral neuropathy; nervous system tumour; neurotoxin; tremor;
 KW neuropathy; acute brain injury; peripheral nerve trauma; human; NOVX;
 KW gene therapy; epilepsy; breast tumour associated protein 47.
 OS Homo sapiens.
 PN US2003027158-A1.
 XX 06-FEB-2003.
 PD 15-OCT-2001; 2001US-00977418.
 PF 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX (CURA-) CURAGEN CORP.
 PA Shimkete RA, Fernandes E, Herrman J, Vernet C;
 PI WPI; 2003-492028/46.
 DR N-PSDB; ACD40305.
 XX New nucleic acid sequence encoding a human breast tumor-associated
 PT protein 47-like polypeptide, useful for treating cardiovascular
 PT disorders, neural disorders, diabetes mellitus and cancers.
 XX Example 3; Page 93; 100pp; English.
 PS The invention relates to a new isolated NOV4 nucleic acid. The nucleic
 CC acid is useful for identifying a compound that binds the nucleic acid.
 CC The nucleic acid is useful in gene therapy, in screening assays, in
 CC detection assays e.g. chromosomal mapping, cell and tissue typing and
 CC forensic biology, predictive medicine e.g. diagnostic assays, prognostic
 CC assays, monitoring clinical trials and pharmacogenomics and methods of
 CC treatment including therapeutic and prophylactic. The nucleic acid is
 CC also useful for expressing NOVX protein. The nucleic acid is also useful
 CC to provide polynucleotide reagents e.g. labelled probes that are useful

CC in an in situ hybridization technique, for identifying a specific tissue
 CC (for example brain tissue) and for use in forensic science. The nucleic
 CC acid is also useful for mapping genes on a chromosome and thus locating
 CC gene regions associated with genetic disease, identifying an individual
 CC from a minute biological sample and to aid in forensic identification of
 CC biological sample. The nucleic acid is also useful for treating cancer,
 CC especially cancers of the breast, colon, lung, pancreas or uterus, or a
 CC melanoma or sarcoma. The nucleic acid is also useful for treating
 CC disorders related to organ transplantation, cardiovascular diseases,
 CC atherosclerosis, ischaemic heart disease, haemorrhage, diabetes mellitus,
 CC peripheral vascular disease, thrombosis, hypertension and systemic lupus
 CC erythematosus. NOVX protein encoded by the nucleic acid is useful for
 CC regulating haematopoiesis, for regeneration of bone, cartilage, tendon
 CC ligament and/or nerve tissue growth or regeneration and for wound
 CC healing. The nucleic acid is also useful for treating infections,
 CC hyperproliferative disorders e.g. psoriasis, and neural disorders,
 CC including Parkinson's disease, Alzheimer's disease, Huntington's disease,
 CC multiple sclerosis, amyotrophic lateral sclerosis (ALS), peripheral
 CC neuropathy, tumours of the nervous system, exposure to neurotoxins, acute
 CC brain injury, peripheral nerve trauma or injury and other neuropathies, acute
 CC epilepsy, and/or tremors. The present sequence represents the amino acid
 CC sequence of a human breast tumour associated protein 47-like polypeptide
 SQ Sequence 101 AA;

Query Match 36.8%; Score 312; DB 6; Length 101;
 Best Local Similarity 60.0%; Pred. No. 5, 4e-26;
 Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

QY 1 MPELYIDGVPFIQVREKLTETPAVRLILGPPGAKWGWGMSFSLPQACVXGAAPMOT 60
 DQ 1 MPELYIDGVPFIQVREKLTETPAVRLILGPPGAKWGWGMSFSLPQACVXGAAPMOT 60
 QY 61 GSPNRDGSQGMG---ALGPRGWTPASCHLPLRQSLALHPSLSLI 101
 DQ 61 DHLTMEKARAWMGLWLVLEVPQ---LLATPSSQSALHPSLSLI 101

RESULT 10
 ID ADF66803 standard; protein, 101 AA.

AC ADF66803;

DT 12-FEB-2004 (first entry)

DE Novel human protein NOV4 gene fragment protein.

XX cytoskeletal; hepatotropic; vulnery; antipsoriatic; osteopathic;
 KM antithyroid; antiatherosclerotic; haemostatic; vasotrophic;
 KM antithyroid; antidiabetic; hypotensive; dermatological;
 KM immunosuppressive; antinflammatory; immunostimulant; fungicide;
 KM virucide; protozoacide; neuroprotective; antineuritic; antiarthritic;
 KM antiasthmatic; antiparkinsonian; nootropic; anticonvulsant;
 KM NOVX modulator; cancer; hyperproliferative disease; cirrhosis; keloid;
 KM psoriasis; tissue hypertrophy; osteoarthritis;
 KM atherosclerotic plaque formation; haemorrhage; ischaemic disease;
 KM thrombosis; diabetes mellitus; hypertension; hypothyroidism;
 KM immune deficiency; severe combined immunodeficiency; SCID; infection;
 KM malaria; candidiasis; autoimmune disorder; connective tissue disease;
 KM multiple sclerosis; systemic lupus, erythematosus; rheumatoid arthritis;
 KM autoimmune pulmonary inflammation; Guillain-Barre syndrome;
 KM autoimmune thyroiditis; insulin dependent diabetes mellitus;
 KM myasthenia gravis; graft-versus-host disease; haematopoiesis;
 KM autoimmune inflammatory eye disease; asthma; haematopoiesis;
 KM tissue regeneration; wound healing; tissue repair; burn; incision; ulcer;
 KM periodontal disease; Alzheimer's disease; Parkinson's disease;
 KM Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
 human.

XX Homo sapiens.
 OS US2003199103-A1.
 XX

XX 23-OCT-2003.
 PD 15-OCT-2001; 2001US-00977639.
 XX 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 PA (CURA-) CUBAGEN CORP.
 PI Shinkets RA, Fernandes E, Herrman J, Vernet C;
 DR WPI; 2004-021196/02.
 DR N-PSDB; ADF66802.
 PT Novel substantially NOVX polypeptide useful for diagnosing, preventing
 PT and treating diseases e.g., cancer, multiple sclerosis, systemic lupus
 PT erythematosus.
 XX Example 3; SEQ ID NO 76; 165pp; English.

XX The invention describes a substantially purified polypeptide (I) having
 CC amino acid sequence chosen from a fully defined NOVX sequence (S1) of 708
 CC amino acids as given in the specification, or polypeptide having one or
 CC more conservative amino acid substitutions of (S1), or mutant or variant
 CC of (S1). (I) Having (S1) is useful for diagnosing a pathological
 CC condition associated with (I) or its activity in a subject e.g. cancer.
 CC (I) useful in treatment of cancer, hyperproliferative diseases,
 CC cirrhosis, keloid, psoriasis, tissue hypertrophy, osteoarthritis,
 CC atherosclerotic plaque formation, haemorrhage, ischaemic heart or renal
 CC disease, thrombosis, diabetes mellitus, hypertension, hypothyroidism. (I)
 CC is useful in treatment of various immune deficiencies and disorders such
 CC as severe combined immunodeficiency (SCID), bacterial infection, viral
 CC infection such as herpes viral infection, protozoan infection such as
 CC malaria, fungal infection such as candidiasis. (I) is also useful in
 CC treating autoimmune disorders such as connective tissue disease, multiple
 CC sclerosis, systemic lupus, erythematosus, rheumatoid arthritis,
 CC autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune
 CC thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis,
 CC graft-versus-host disease and autoimmune inflammatory eye disease and
 CC asthma. (I) useful in regulation of haematopoiesis, regeneration and
 CC tissue growth of bone, cartilage, tendon, ligament and useful for wound
 CC healing and tissue repair. (I) is also useful in treatment of burns,
 CC incisions and ulcers. (I) also useful in treatment of periodontal
 CC disease, Alzheimer's disease, Parkinson's disease, Huntington's disease,
 CC amyotrophic lateral sclerosis, and Shy-Drager syndrome. (I) has effective
 CC antitumour and antinflammatory activity. This is the amino acid sequence
 CC of a novel human NOVX protein.

SQ Sequence 101 AA;

Query Match 36.8%; Score 312; DB 6; Length 101;
 Best Local Similarity 60.0%; Pred. No. 5, 4e-26;
 Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

QY 1 MPELYIDGVPFIQVREKLTETPAVRLILGPPGAKWGWGMSFSLPQACVXGAAPMOT 60
 DQ 1 MPELYIDGVPFIQVREKLTETPAVRLILGPPGAKWGWGMSFSLPQACVXGAAPMOT 60
 QY 61 GSPNRDGSQGMG---ALGPRGWTPASCHLPLRQSLALHPSLSLI 101
 DQ 61 DHLTMEKARAWMGLWLVLEVPQ---LLATPSSQSALHPSLSLI 101

RESULT 11
 ID ADO60313 standard; protein, 101 AA.
 XX ADO60313;
 AC

XX 15-JUL-2004 (first entry)
 XX Human NOVA protein fragment.
 DE
 XX Human, NOVA protein; cancer; hyperproliferative disease; cirrhosis;
 KW Keloid; psoriasis; tissue hypertrophy; osteoarthritis;
 KW atherosclerotic plaque formation; haemorrhage; ischaemic heart disease;
 KW renal disease; thrombosis; diabetes mellitus; hypertension;
 KW hypothyroidism; severe combined immunodeficiency; SCID; infection;
 KW malaria; candidiasis; autoimmune disorder; connective tissue disease;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW autoimmune pulmonary inflammation; Guillain-Barre syndrome;
 KW autoimmune thyroiditis; myasthenia gravis; graft-versus-host disease;
 KW autoimmune inflammatory eye disease; asthma; burn; incision; ulcer;
 KW pericardial disease; Alzheimer's disease; Parkinson's disease;
 KW Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
 KW haematopoiesis; wound healing; tissue repair; antitumor;
 KW antiinflammatory.
 KW
 XX Homo sapiens.
 OS
 XX US2003134430-A1.
 PN
 XX 17-JUL-2003.
 PD
 XX 15-OCT-2001; 2001US-00977751.
 PE
 XX 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 XX Shimkets RA, Fernandes E, Herrman J, Vernet C;
 PI WPI: 2004-068928/07.
 DR N-PSDB; AD060312.
 XX
 XX Novel substantially purified NOVA polypeptide for treating severe
 PT combined immunodeficiency, candidiasis, cancer, asthma, multiple
 PT sclerosis, systemic lupus erythematosus.
 PT
 XX Example 3; SEQ ID NO 76; 155dp; English.
 PS
 XX The invention relates to human NOVA polypeptides and polynucleotides.
 CC NOVA sequences are useful in the treatment of cancer, hyperproliferative
 CC diseases, cirrhosis, keloid, psoriasis, tissue hypertrophy,
 CC osteoarthritis, atherosclerotic plaque formation, haemorrhage, ischaemic
 CC heart or renal disease, thrombosis, diabetes mellitus, hypertension,
 CC hypothyroidism, asthma, burns, incisions, ulcers, periodontal disease,
 CC Alzheimer's disease, Parkinson's disease, Huntington's disease,
 CC amyotrophic lateral sclerosis, Shy-Drager syndrome, immune deficiencies
 CC and disorders such as severe combined immunodeficiency (SCID), bacterial
 CC infection, viral infection e.g. herpes viral infection, protozoan
 CC infection e.g. malaria, fungal infection e.g. candidiasis, autoimmune
 CC disorders such as connective tissue disease, multiple sclerosis, systemic
 CC lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
 CC inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin
 CC dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease
 CC and autoimmune inflammatory eye disease. The invention is useful in
 CC regeneration and tissue growth of bone, cartilage, tendon, ligament,
 CC haematopoiesis regulation, wound healing and tissue repair. Sequences of
 CC the invention also exhibit antitumor and antiinflammatory activities. The
 CC present sequence is human NOVA protein fragment used in the
 CC exemplification of the invention.
 CC
 XX Sequence 101 AA;
 SQ

Query Match 36.8%; Score 312; DB 8; Length 101;

Best Local Similarity 60.0%; Pred. No. 5,4e-26;
 Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;
 QY 1 MPELYIIOGVFPICIVREKTEETPAVLLILGPPGAKKMGWGSFSLPQACVYGAARMQT 60
 DB 1 MPELYIIOGVFPICIVREKTEETPAVLLILGPPGAKKMGWGSFSLPQACVYGAARMQT 60
 QY 61 GSPNRDQSGMDG---ALGPRGTPASCHLPLRQSLHPSPLI 101
 DB 61 DHELEMEARAWMGJMWLEVGPO---LLATPPSQSLHPSPLI 101
 RESULT 12
 ADM56440
 ID ADM56440 standard; protein; 101 AA.
 XX
 AC ADM56440;
 XX
 DT 03-JUN-2004 (first entry)
 DE
 XX Human cell adhesion molecule NOVA #2.
 XX
 XX Human; cell adhesion molecule; NOVA; cancer; leukaemia; lymphoma;
 KW melanoma; neurological disorder; epilepsy;
 KW ischaemic cerebrovascular disease; stroke; Alzheimer's disease;
 KW Pick's disease; vesicular transport disorder; cystic fibrosis;
 KW diabetes mellitus; Grave's disease; goiter; gastrointestinal disorder;
 KW ulcerative colitis; gastric ulcer; duodenal disorder; autoimmune disease;
 KW allergic reaction; autoimmune haemolytic anaemia; rheumatoid arthritis;
 KW viral infection; bacterial infection; fungal infection;
 KW helminthic infection; protozoal infections.
 KW
 XX Homo sapiens.
 OS
 XX US2003082554-A1.
 PN
 XX 01-MAY-2003.
 PD
 XX 15-OCT-2001; 2001US-00977703.
 PE
 XX 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 XX Shimkets RA, Fernandes E, Herrman J, Vernet C;
 PI WPI: 2003-616079/58.
 DR N-PSDB; ADM56439.
 XX
 XX New nucleic acids encoding human cell adhesion molecule-like proteins,
 PT useful for treating e.g. cancer, neurological disorders, viral,
 PT bacterial, fungal, helminthic and protozoal infections.
 PT
 XX Example 3; SEQ ID NO 76; 78dp; English.
 PS
 XX The invention relates to an isolated nucleic acid encoding a human cell
 CC adhesion molecule-like protein, comprising a sequence encoding a
 CC polypeptide having a sequence appearing as ADM56387, a sequence at least
 CC 90% identical to the nucleic acid, a sequence encoding a polypeptide
 CC having conservative amino acid substitutions to the protein or a fragment
 CC comprising at least 20 nucleotides. Also included are an oligonucleotide
 CC sequence that is complementary to (and hybridises under stringent
 CC conditions with) the nucleic acid (or a portion of it), a vector
 CC comprising the nucleic acid, a cell comprising the vector, a
 CC pharmaceutical composition comprising the nucleic acid and a
 CC pharmaceutical carrier, a process for producing the polypeptide, a
 CC process for identifying a compound that binds the nucleic acid, and a
 CC compound identified by the process. Disclosed as new are the cDNA and

CC proteins for novel cell adhesion molecules (termed NOVX, being NOV1-23).
 CC The NOVX polypeptide, nucleic acid or antibody are useful in the
 CC manufacture of a medicament for treating a syndrome associated with a
 CC human disease selected from NOVX-associated disorder, such as cancers
 CC (e.g. leukaemia, lymphoma, melanoma or cancer of the liver, lung, muscle,
 CC ovary, testis and uterus), neurological disorders (e.g. epilepsy, Pick's
 CC ischaemic cerebrovascular disease, stroke, Alzheimer's disease or Pick's
 CC disease), disorders of vesicular transport (e.g. cystic fibrosis,
 CC diabetes mellitus, Grave's disease, or goiter), gastrointestinal
 CC disorders (e.g. ulcerative colitis, or gastric and duodenal disorders),
 CC autoimmune diseases (e.g. allergic reactions, autoimmune haemolytic
 CC anaemia, or rheumatoid arthritis), viral, bacterial, fungal, helminthic
 CC and protozoal infections. The polypeptides can be used as immunogens to
 CC produce antibodies and as vaccines. The sequences may further be used in
 CC chromosome mapping, identifying individual from minute biological samples
 CC (tissue typing), and in forensic identification of a biological sample.
 CC NOTE: The authors have mislabelled the sequences as they appear on pages
 CC 12-28 of the patent, it is clear from table 3, the examples and the
 CC claims that the SEQ ID numbers for the cDNAs should be the odd numbers
 CC from 1-45 and the proteins should be the even numbers from 2-46. The
 CC present sequence represents a NOVX cell adhesion molecule of the
 CC invention.

XX Sequence 101 AA;

Query Match 36.0%; Score 305; DB 7; Length 101;
 Best Local Similarity 59.0%; Pred. No. 3.2e-25;
 Matches 62; Conservative 7; Mismatches 28; Indels 8; Gaps 2;

QY 1 MPEHYIGVFPFIQVREKIHETPAVRLILGPRGKAKWGMGMSFSLPQACVXGAAPMOT 60
 DB 1 MPEHYIGVFPFIQVREKIHETPAVRLILGPRGKAKWGMGMSFSLPQACVXGAAPMOT 60

QY 61 GSENRDSQGMDS---ALGPRGWTASCHLPLRQSALHPSLSLI 101
 DB 61 DHLTMEARAWMGVLWLEVGPO---LILTFPSGQSALHPSLSLI 101

RESULT 13

ID AD119840 standard; protein; 101 AA.

XX AD119840;

DT 22-APR-2004 (first entry)

XX Human NOV4 protein fragment.

XX Secreted protein; NOVX; diagnosis; metabolic disorder; diabetes; obesity;
 KW infection; anorexia; cancer; cardiovascular disease; hypertension;
 KW atherosclerosis; neurodegenerative disorder; Alzheimer's disease;
 KW Parkinson's disease; epilepsy; immune disorder; osteoarthritis;
 KW haematopoietic disorder; inflammatory skin disorder; asthma;
 KW dyslipidemia; neurogenesis; cell differentiation; cell proliferation;
 KW haematopoiesis; wound healing; angiogenesis; chromosome mapping;
 KW tissue typing; preventive medicine; pharmacogenomic; gene therapy;
 KW anorectic; cardiant; virucide; antibacterial; fungicide; protozoacide;
 KW nootropic; neuroprotective; dermatological; human.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 27 /note= "Encoded by CGC"

FT Misc-difference 40 /note= "Encoded by GCG"

FT Misc-difference 51 /note= "Encoded by CTG"

XX US2004002134-A1.

XX 01-JAN-2004.

PF 15-OCT-2001; 2001US-00977819.

XX 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.

XX (CURA-) CURAGEN CORP.

PI Shimetsu RA, Fernandes ER, Herrman JL, Vernet CM;

DR WPI; 2004-070737/07.

XX N-PSDB; AD119839.

PT New NOVX nucleic acids encoding human KIA0768 protein-like and human
 PT protein PRO-228 polypeptides, useful for treating NOVX-associated
 PT disorders.

XX Example 3; SEQ ID NO 76; 95PP; English.

CC The present invention is based in part on the discovery of novel secreted
 CC and membrane-bound polypeptides and their encoding polynucleotides. The
 CC nucleic acids and polypeptides are collectively referred as NOVX. The
 CC invention is useful for treating, preventing and diagnosing diseases such
 CC as metabolic disorders, diabetes, obesity, infectious diseases such as
 CC viral, bacterial, fungal, helminthic and protozoal infections, anorexia,
 CC cancer, cardiovascular diseases such as hypertension and atherosclerosis,
 CC neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,
 CC epilepsy, immune disorders such as osteoarthritis, haematopoietic
 CC disorders, inflammatory skin disorders, asthma and various dyslipidemias.
 CC The invention is also useful as targets for the identification of small
 CC molecules that modulate or inhibit e.g. neurogenesis, cell
 CC differentiation, cell proliferation, haematopoiesis, wound healing and
 CC angiogenesis, as hybridisation probes, in chromosome mapping, tissue
 CC typing, preventive medicine and pharmacogenomics. The invention is also
 CC useful in gene therapy. The present sequence is human NOV protein
 CC fragment.

XX Sequence 101 AA;

Query Match 34.9%; Score 296; DB 8; Length 101;
 Best Local Similarity 58.1%; Pred. No. 3.1e-24;
 Matches 61; Conservative 7; Mismatches 29; Indels 8; Gaps 2;

QY 1 MPEHYIGVFPFIQVREKIHETPAVRLILGPRGKAKWGMGMSFSLPQACVXGAAPMOT 60
 DB 1 MPEHYIGVFPFIQVREKIHETPAVRLILGPRGKAKWGMGMSFSLPQACVXGAAPMOT 60

QY 61 GSENRDSQGMDS---ALGPRGWTASCHLPLRQSALHPSLSLI 101
 DB 61 DHLTMEARAWMGVLWLEVGPO---LILTFPSGQSALHPSLSLI 101

RESULT 14

ID ABB05728 standard; protein; 703 AA.

XX ABB05728;

DT 30-APR-2002 (first entry)

XX Human transmembrane protein clone tes3_7n12.

XX Human; foetal brain; foetal kidney; melanoma; testis; amygdala;

XX gene therapy.

XX Homo sapiens.

XX WO200198454-A2.

XX 27-DEC-2001.

```

XX 25-APR-2001; 2001MO-IB002050.
XX
XX 25-APR-2000; 2000US-0199380P.
XX
XX (GERH-) GERMAN HUMAN GENOME PROJECT.
XX
XX Wiemann S;
XX
XX WPI: 2003-055860/07.
XX
XX N-PSDB; ABA93765.
XX
XX Human cDNA sequences and clones derived from human fetal brain, fetal
XX kidney, melanoma, testis and amygdala cDNA libraries, useful in genetic
XX screening and therapy.
XX
XX Claim 1; Page 386; 611pp; English.
XX
XX The present invention describes assemblies and computer readable media
XX comprising novel human cDNA sequences and clones derived from human
XX foetal brain, foetal kidney, melanoma, testis and amygdala cDNA
XX libraries. ABA93702 to ABA93766 represent human cDNA sequences from the
XX present invention which encode the proteins given in ABA05662 to
XX ABA05729. The human cDNA sequences and clones can be used in gene
XX therapy. The clones may be used in a variety of applications, for example
XX they may be used in profiling assays, for providing large arrays of human
XX genetic material for implementing large-scale screening strategies and
XX for treating diseases via gene therapy procedures
XX
XX Sequence 703 AA;
XX
XX Query Match 13.4%; Score 113.5; DB 5; Length 703;
XX Best Local Similarity 25.3%; Pred. No. 0.0035;
XX Matches 45; Conservative 20; Mismatches 60; Indels 53; Gaps 9;
XX
XX 20 HETPAVRLT--GPPGKAKWGMCM---GWSFSLPOACYGAAPMOTGSPNRDGSQMDG 73
XX Db 214 HGTLIGDLVDAGAGTASGCTKDLAPWVNL--LCHSMACPMSCGPRPRQUTVD 271
XX QY 74 A-----LQPR-GWTPAS-----CHLPKQSLHPS- 98
XX Db 272 ARLHRELSLVPRIGVLAATPMRTWGPFGFPFASIRARCTCHVCHRSFPAKLTFCPCQC 331
XX QY 99 -SLIYESIGSVCTTHPYRCPSNTOHYHM-----LRNSELSCIRFTISFINGET 148
XX Db 332 SAVLY--GPRACLRADWQRCDDVSHRFMCPRLAAMERAGELATLPFTYIAEVISET 387
XX
XX RESULT 15
XX ADB64988
XX ID ADB64988 standard; protein; 247 AA.
XX
XX AC ADB64988;
XX
XX DT 04-DEC-2003 (first entry)
XX
XX DE Human protein encoded by clone PROST20122490.
XX
XX KW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
XX cell regeneration; membrane protein; signal transduction-related protein;
XX transfection-related protein; osteoporosis; neurological disease;
XX cancer; tumour.
XX
XX OS Homo sapiens.
XX
XX PN EP1308459-A2.
XX
XX PD 07-MAY-2003.
XX
XX PF 28-MAR-2002; 2002EP-00007401.
XX
XX PR 05-NOV-2001; 2001JP-00379298.
XX
XX 25-JAN-2002; 2002US-00350978.

```

```

XX (HELI-) HELIX RES INST.
XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Iisogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;
XX
XX WPI: 2003-450961/43.
XX
XX N-PSDB; ADB63018.
XX
XX New polynucleotides and polypeptides, useful for developing a diagnostic
XX marker or medicines for regulation of their expression and activity, or
XX as targets of gene therapy.
XX
XX Claim 1; Page: 222pp; English.
XX
XX The invention discloses a polynucleotide comprising a sequence selected
XX from 1970 fully defined nucleotide sequences which encode novel
XX polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
XX or its partial peptide, an antibody binding to the polypeptide or peptide
XX of the polynucleotide, immunologically assaying the polypeptide or
XX peptide of the polynucleotide by contacting the polypeptide or peptide
XX with the antibody of the encoded protein, and observing the binding
XX between the two, a transformant carrying the polynucleotide in an
XX expressible manner and an antisense polynucleotide. The oligonucleotide
XX is useful as a primer for synthesizing the polynucleotide, or as a probe
XX for detecting the polynucleotide. The polynucleotides and encoded
XX proteins are useful as pharmaceutical agents and many disease-related
XX genes may be included in them, for developing a diagnostic marker or
XX medicines for regulation of their expression and activity, or as targets
XX of gene therapy. The genes are involved in tissue and/or cell
XX regeneration. Membrane proteins, signal transduction-related proteins,
XX transcription-related proteins, disease-related proteins and genes
XX encoding them can be used as indicators for diseases (e.g. osteoporosis,
XX neurological diseases, cancer, tumours. The cDNA may be used to regulate
XX the activity or expression of the encoded protein to treat diseases. The
XX sequence presented is a protein of the invention. Note: Some of the
XX sequence data for this patent is not represented in the printed
XX specification, but is based on sequence information supplied by the
XX European Patent Office.
XX
XX Sequence 247 AA;
XX
XX Query Match 11.2%; Score 95; DB 7; Length 247;
XX Best Local Similarity 25.9%; Pred. No. 0.096;
XX Matches 35; Conservative 11; Mismatches 29; Indels 60; Gaps 7;
XX
XX 2 PHLIYDGFPIQVREKLA--ETPAV-----LRLGPPGK-----AKW----- 37
XX Db 105 PHRYVAHLVYARQLTEGLHVAEISAVHSRHHGVEVGPBPRPPRGLALAPKAPWGMT 164
XX QY 38 -----GM-----GWSFSLPOACYGAAPMOTGSPNR 65
XX Db 165 SPEGPPLAGLRVSLSPAMGWEGSGAGRPKEGRGWPGLDLPQGVTSALPLAIANP-- 222
XX QY 66 DGSQMDGALGPRGM 80
XX Db 223 -GSGGV-GSVGRKGM 235

```

Search completed: October 24, 2004, 16:01:44
Job time : 80 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2004, 15:44:07 ; Search time 24 Seconds
(without alignments)
420.014 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848
Sequence: 1 MPRHYDGVFPIQLVREKH.....LSQIRPTISIFINGRTSPSH 152

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PTCUS COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	10.4	1745	4	US-09-795-061-4 Sequence 4, Appl 1
2	84.5	10.0	334	4	US-09-252-991A-16629 Sequence 16629, A
3	82.5	9.7	884	2	US-08-463-976A-2 Sequence 2, Appl 1
4	82.5	9.7	884	2	US-08-982-412-2 Sequence 2, Appl 1
5	81.5	9.6	457	4	US-09-252-991A-29671 Sequence 29671, A
6	81	9.6	1739	4	US-09-795-061-2 Sequence 2, Appl 1
7	80	9.4	185	2	US-08-463-911-3 Sequence 3, Appl 1
8	80	9.4	365	4	US-09-489-847-322 Sequence 322, App
9	80	9.4	542	4	US-09-489-847-323 Sequence 323, App
10	79	9.3	215	4	US-09-511-176B-51 Sequence 51, Appl
11	79	9.3	215	4	US-09-619-740-54 Sequence 54, Appl
12	79	9.3	236	3	US-09-140-804-6 Sequence 6, Appl 1
13	79	9.3	236	4	US-09-686-838B-6 Sequence 6, Appl 1
14	79	9.3	272	4	US-09-252-991A-27652 Sequence 27652, A
15	78	9.2	75	4	US-09-513-999C-5144 Sequence 5144, Ap
16	77.5	9.1	178	4	US-09-252-991A-23713 Sequence 23713, A
17	77.5	9.1	300	2	US-08-712-948-2 Sequence 2, Appl 1
18	77.5	9.1	405	4	US-08-755-235-4 Sequence 4, Appl 1
19	77	9.1	318	2	US-08-633-148-4 Sequence 4, Appl 1
20	77	9.1	332	4	US-09-062-365-1 Sequence 1, Appl 1
21	77	9.1	340	2	US-08-633-148-2 Sequence 2, Appl 1
22	77	9.1	404	4	US-09-638-649-3 Sequence 3, Appl 1
23	77	9.1	1086	4	US-09-252-991A-25051 Sequence 25051, A
24	76	9.0	733	2	US-08-712-948-1 Sequence 1, Appl 1
25	76	9.0	772	4	US-09-404-879A-388 Sequence 388, App
26	76	9.0	772	4	US-09-667-857-388 Sequence 388, App
27	76	9.0	914	4	US-09-404-879A-312 Sequence 312, App

28	76	9.0	914	4	US-09-338-933-312 Sequence 312, App
29	76	9.0	914	4	US-09-667-857-312 Sequence 312, App
30	75.5	8.9	498	5	PCR-US94-01101-2 Sequence 2, Appl 1
31	75	8.8	593	4	US-09-252-991A-20441 Sequence 20441, A
32	75	8.8	1442	2	US-08-316-650-12 Sequence 12, Appl 1
33	75	8.8	1442	5	PCR-US95-02251-12 Sequence 12, Appl 1
34	74	8.7	391	4	US-09-270-767-58641 Sequence 58641, A
35	74	8.7	608	4	US-09-270-767-43297 Sequence 43297, A
36	74	8.7	1184	1	US-08-446-038B-20 Sequence 20, Appl 1
37	74	8.7	1184	1	US-08-446-010B-20 Sequence 20, Appl 1
38	74	8.7	1184	2	US-08-805-445-20 Sequence 20, Appl 1
39	74	8.7	1184	2	US-08-064-067D-20 Sequence 20, Appl 1
40	74	8.7	1184	2	US-09-066-208-20 Sequence 20, Appl 1
41	74	8.7	1187	1	US-08-357-598-8 Sequence 8, Appl 1
42	74	8.7	1187	1	US-08-097-997A-13 Sequence 13, Appl 1
43	74	8.7	1187	2	US-09-003-289-8 Sequence 8, Appl 1
44	74	8.7	1187	3	US-08-665-574C-13 Sequence 13, Appl 1
45	74	8.7	1187	3	US-08-946-994-13 Sequence 13, Appl 1

ALIGNMENTS

RESULT 1
US-09-795-061-4
; Sequence 4, Application US/09795061
; Patent No. 6759528
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S
; TITLE OF INVENTION: Imamura, Yasutada
; FILE REFERENCE: 960296.96781
; CURRENT APPLICATION NUMBER: US/09/795,061
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1745
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-795-061-4

Query Match 10.4%; Score 88; DB 4; Length 1745;
Best Local Similarity 24.2%; Pred. No. 1;

Matches 45; Conservative 17; Mismatches 54; Indels 70; Gaps 9;

QY	25	VRILGPPGKAKWGMGMSFSLPQACVYG-----AAPMGTGSPN	64
DB	1407	LIGLIPPEEA-----GEKDGQILPG--VGGPPGKGDPPGPGTIGLHGPSPVAGPLG	1460
QY	65	RDSQGMGALGPRGWT-----PASCHPLRSGALHPSFLIVE-----SI	105
DB	1461	QKSKSKSPSMGPRGDTGAGPPGPGAPAEHLHGRRRRFFVPLPVVEGGLSEYIASL	1520
QY	106	GSVCVTHRYRCPSNT-----QHYN-----WLPNSELGIRTTISFTN-	145
DB	1521	TSLSLELDLRPPGTAERPGLVCHLHRNHPHLPDGEYWDPNQCA--RDSFRVCFNF	1578
QY	146	---GET 148	
DB	1579	TARGET 1584	

RESULT 2
US-09-252-991A-16629
; Sequence 16629, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 16629
LENGTH: 334
TYPE: PR1
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16629

Query Match 10.0%; Score 84.5; DB 4; Length 334;
Best Local Similarity 24.4%; Pred. No. 0.3;
Matches 38; Conservative 17; Mismatches 46; Indels 55; Gaps 11;

QY 21 ETFAVRL-LGPRGXKMG-----MGWG-----SPSLPQACVXGAAPQGTSP 63
DB 50 ETVVWGTRFGALG-GSMGPKFPGRVPGWGMGRRASWPRCTGAVSP----- 102
QY 64 NRDGSGQMDGAL---GPRGW---TPASCHLPLRGSALHPSPLYESIGS---VCVTT 112
DB 103 -GGRMPDRCAGFQAGPPPRAWRSPGT-----SPALVAPRGAAPRIVAVTA 148
QY 113 HPRCPSTNTOHYHMLAPNSLS--QIRFTISIFING 146
DB 149 ---CPTGVAHPTMAAEALQLAGQLGALQVETQG 180

RESULT 3

US-08-465-976A-2
Sequence 2, Application US/08465976A
Patent No. 5869632
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R
APPLICANT: LI, YI
APPLICANT: ROSEN, CRAIG A
APPLICANT: RUBEN, STEVEN M
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN GILFILLAN, CECCHI
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NJ
COUNTRY: US
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465, 976A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY F
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-444
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 994-1700
TELEFAX: (201) 994-1744
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-465-976A-2

Query Match 9.7%; Score 82.5; DB 2; Length 884;

Best Local Similarity 28.9%; Pred. No. 1.8;
Matches 48; Conservative 9; Mismatches 54; Indels 55; Gaps 9;

QY 13 QLVREKIHETPA-VLRLLGPPGKAKMGWGMGSESLPQACVXGA-----APMOTGS----- 62
DB 641 QLSREERLEAPAPVLRPLSRPS-----QECMDAPGRLEPKRSGSTLPRR 686
QY 63 -PNRDGSGQMDGALG-----PRGW-----TPASCH-LPLRGSALHP 96
DB 687 QPRPDYPGAAAGRGSDALDLGAPREWLSTLPPRRTRDLDSPHLCPLPSGNSQGT 746
QY 97 S-PSLIYESIGSVCTTHPRCPSTNTOHYHML--RNSLSQIR 137
DB 747 SCHGRWTLICGARTLGSSWTRCIAGTPHEKPLGHSRSCSELGRTR 792

RESULT 4

US-08-982-412-2
Sequence 2, Application US/08982412
Patent No. 5958729
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R
APPLICANT: LI, YI
APPLICANT: ROSEN, CRAIG A
APPLICANT: RUBEN, STEVEN M
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUMAN GENOME SCIENCES, INC.
STREET: 9410 KEY WEST AVENUE
CITY: ROCKVILLE,
STATE: MD
COUNTRY: US
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/982,412
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: BROOKES, ANDERS A
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PFI91PCT2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8439
TELEFAX: (301) 309-8504
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-982-412-2

Query Match 9.7%; Score 82.5; DB 2; Length 884;
Best Local Similarity 28.9%; Pred. No. 1.8;
Matches 48; Conservative 9; Mismatches 54; Indels 55; Gaps 9;

QY 13 QLVREKIHETPA-VLRLLGPPGKAKMGWGMGSESLPQACVXGA-----APMOTGS----- 62
DB 641 QLSREERLEAPAPVLRPLSRPS-----QECMDAPGRLEPKRSGSTLPRR 686
QY 63 -PNRDGSGQMDGALG-----PRGW-----TPASCH-LPLRGSALHP 96
DB 687 QPRPDYPGAAAGRGSDALDLGAPREWLSTLPPRRTRDLDSPHLCPLPSGNSQGT 746
QY 97 S-PSLIYESIGSVCTTHPRCPSTNTOHYHML--RNSLSQIR 137
DB 747 SCHGRWTLICGARTLGSSWTRCIAGTPHEKPLGHSRSCSELGRTR 792

```
RESULT 5
US-09-252-991A-29671
; Sequence 29671, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29671
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29671

Query Match.          9.6%; Score 81.5; DB 4; Length 457;
Best Local Similarity 33.0%; Pred. No. 0.96;
Matches 33; Conservative 6; Mismatches 44; Indels 17; Gaps 4;

QY 2 PHLVDGVFPQLVREKLEHTPAVLRL-----GPGKAKWGMGMSFSLPQACVYGAA 56
DB 290 PALRRGRIP---ERRPRRLGRLRLRPVGHPPASAD-----RFPAPRQDPGR 339
QY 57 PMQTGSPNEDSGQMDGALGPGWTPASCHLPLRQSLHP 96
DB 340 PRAGAALRRGPR--MGSLARPGRTPAARHPLRDLRHP 377

RESULT 6
US-09-795-061-2
; Sequence 2, Application US/09795061
; Patent No. 6759528
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S
; TITLE OF INVENTION: Pro-Alpha 3 (V) Collagen Genes
; FILE REFERENCE: 960296.96781
; CURRENT APPLICATION NUMBER: US/09/795,061
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-795-061-2

Query Match          9.6%; Score 81; DB 4; Length 1739;
Best Local Similarity 24.2%; Pred. No. 6.2;
Matches 44; Conservative 19; Mismatches 53; Indels 66; Gaps 10;

QY 25 VLRLLGPPGKAKWGMGMSFSLPQACVYGAAPNQ-----TSSPN 64
DB 1406 LIGLIGRPGA-----GEKDGQDLGG--VQGPBGQDGRGLGPGVSLGHPGPPGVGRLG 1459
QY 65 RDGSGCMGALGPRG-----WTPASCH-LPLRQS---ALHSPSLIYESIGSVC 109
DB 1460 QKSGSGSGSLGPRDPPGAPPPGPGSPALVHGLRRRSVDTLLEGLEBVMASLNSLS 1519
QY 110 VTHHPYRCPSNT-----QHYH-----WLRPNSELQIFFTTISIFIN---G 146
DB 1550 LELQQLGRLPTAESPGLMCRELHRDHPLDGBEYWDIPNGCARDAF--KVFENFTAGG 1577
QY 147 ET 148
```

```
DB 1578 ET 1579

RESULT 7
US-08-463-911-3
; Sequence 3, Application US/08463911
; Patent No. 5869330
; GENERAL INFORMATION:
; APPLICANT: Scherer, Philipp E.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: A NOVEL SERUM PROTEIN PRODUCED
; TITLE OF INVENTION: EXCLUSIVELY IN ADIPOCYTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,911
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI95-05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 185 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-463-911-3

Query Match          9.4%; Score 80; DB 2; Length 185;
Best Local Similarity 23.5%; Pred. No. 0.44;
Matches 23; Conservative 13; Mismatches 32; Indels 30; Gaps 4;

QY 48 PQACVYGAAPMQTGSNRRGSGMDGALGPRGWT-----PASCH-----L 87
DB 6 PESC--NAPGQGPQGPGPPGPPGPMNGFPJGPPGPGKMTVNCHSKGTSAF 62
QY 88 PLRQSLHPSPS-----LIYESIGSVCTTHPYRCP 118
DB 63 AYVANELPPAPGQPVIFKEALHDAQHPDLATGVFTCP 100

RESULT 8
US-09-489-847-322
; Sequence 322, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: P2031P1
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
```

EARLIER APPLICATION NUMBER: 60/095,486
EARLIER FILING DATE: 1998-08-05
EARLIER APPLICATION NUMBER: 60/096,319
EARLIER FILING DATE: 1998-08-12
EARLIER APPLICATION NUMBER: 60/095,454
EARLIER FILING DATE: 1998-08-06
EARLIER APPLICATION NUMBER: 60/095,455
EARLIER FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 376
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 322
LENGTH: 365
TYPE: PRT
ORGANISM: Homo sapiens
US-09-489-847-322

Query Match 9.4%; Score 80; DB 4; Length 365;
Best Local Similarity 25.9%; Pred. No. 1.1;
Matches 30; Conservative 12; Mismatches 30; Indels 44; Gaps 4;

12 IGVREKLE-----TPAVRLILGPGKAKMGWGMGWSFSLPQACVXGAAP 57
Db 219 LQLEBRLHQLSLKDLTGPAEGAPPGPGLQGPPEPAG----- 257
QY 58 MGTGSPNRDGSQGMGALGPRG---WTPASCHLPLRQSALHPSPSLIYESIGSV 108
Db 258 -PGSGKXGQGGEPGIPGPBGQGGVGEAPAA---PVPQVAFSAALSLPRSEPGTV 309

RESULT 9
US-09-489-847-323

Sequence 323, Application US/09489847
Patent No. 6476195
GENERAL INFORMATION:
APPLICANT: Rosen et al
TITLE OF INVENTION: 98 Human Secreted Proteins
FILE REFERENCE: P2031P1
CURRENT APPLICATION NUMBER: US/09/489,847
CURRENT FILING DATE: 2000-01-24
EARLIER APPLICATION NUMBER: PCT/US99/17130
EARLIER FILING DATE: 1999-07-29
EARLIER APPLICATION NUMBER: 60/094,657
EARLIER FILING DATE: 1998-07-30
EARLIER APPLICATION NUMBER: 60/095,486
EARLIER FILING DATE: 1998-08-05
EARLIER APPLICATION NUMBER: 60/096,319
EARLIER FILING DATE: 1998-08-12
EARLIER APPLICATION NUMBER: 60/095,454
EARLIER FILING DATE: 1998-08-06
EARLIER APPLICATION NUMBER: 60/095,455
EARLIER FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 376
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 323
LENGTH: 542
TYPE: PRT
ORGANISM: Homo sapiens
US-09-489-847-323

Query Match 9.4%; Score 80; DB 4; Length 542;
Best Local Similarity 25.9%; Pred. No. 1.8;
Matches 30; Conservative 12; Mismatches 30; Indels 44; Gaps 4;

12 IGVREKLE-----TPAVRLILGPGKAKMGWGMGWSFSLPQACVXGAAP 57
Db 323 LQLEBRLHQLSLKDLTGPAEGAPPGPGLQGPPEPAG----- 361
QY 58 MGTGSPNRDGSQGMGALGPRG---WTPASCHLPLRQSALHPSPSLIYESIGSV 108
Db 362 -PGSGKXGQGGEPGIPGPBGQGGVGEAPAA---PVPQVAFSAALSLPRSEPGTV 413

RESULT 10

US-09-911-176B-51
Sequence 51, Application US/09911176B
Patent No. 6518403
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
TITLE OF INVENTION: ANTIBODIES THAT BIND AN
TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOG
FILE REFERENCE: 97-30D1
CURRENT APPLICATION NUMBER: US/09/911,176B
CURRENT FILING DATE: 2001-07-23
PRIOR APPLICATION NUMBER: 09/118,408
PRIOR FILING DATE: 1998-07-17
PRIOR APPLICATION NUMBER: 60/053,154
PRIOR FILING DATE: 1997-07-18
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 51
LENGTH: 215
TYPE: PRT
ORGANISM: Tamias sibiricus
US-09-911-176B-51

Query Match 9.3%; Score 79; DB 4; Length 215;
Best Local Similarity 23.5%; Pred. No. 0.69;
Matches 24; Conservative 14; Mismatches 28; Indels 36; Gaps 4;

30 GPGKAKMGWGMGWSFSLPQACVXGAAPMOTGSPNRDGSQGMGALGPRGWTPTASCH--- 86
Db 52 GPGG-----TPGKPGPPGNNGPPGLPBPBPBGMT-VNCHSKG 88
QY 87 ---LPLRQSALHPSPS-----LIYESIGSVCTTHPRCP 118
Db 89 TSAFAVAVANLPPAPSGPVITFKEALHDAQHPDLATGVFTCP 130

RESULT 11

US-09-619-740-54
Sequence 54, Application US/09619740
Patent No. 6544946
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Lasser, Gerald W.
APPLICANT: Bishop, Paul D
TITLE OF INVENTION: INHIBITORS FOR USE IN HEMOSTASIS AND IMMUNE FUNCTION
FILE REFERENCE: 99-12C3
CURRENT APPLICATION NUMBER: US/09/619,740
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/253,604
PRIOR FILING DATE: 1999-02-19
PRIOR APPLICATION NUMBER: 09/444,794
PRIOR FILING DATE: 1999-11-22
PRIOR APPLICATION NUMBER: 09/506,855
PRIOR FILING DATE: 2000-02-17
NUMBER OF SEQ ID NOS: 55
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 54
LENGTH: 215
TYPE: PRT
ORGANISM: Tamias sibiricus
US-09-619-740-54

Query Match 9.3%; Score 79; DB 4; Length 215;
Best Local Similarity 23.5%; Pred. No. 0.69;
Matches 24; Conservative 14; Mismatches 28; Indels 36; Gaps 4;

30 GPGKAKMGWGMGWSFSLPQACVXGAAPMOTGSPNRDGSQGMGALGPRGWTPTASCH--- 86
Db 52 GPGG-----TPGKPGPPGNNGPPGLPBPBPBGMT-VNCHSKG 88
QY 87 ---LPLRQSALHPSPS-----LIYESIGSVCTTHPRCP 118
Db 89 TSAFAVAVANLPPAPSGPVITFKEALHDAQHPDLATGVFTCP 130

RESULT 12
US-09-140-804-6
Sequence 6, Application US/09140804
Patent No. 6197930
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
FILE REFERENCE: 97-49
CURRENT APPLICATION NUMBER: US/09/140,804
PRIOR FILING DATE: 1998-08-26
EARLIER APPLICATION NUMBER: 60/056,983
PRIOR FILING DATE: 1997-08-26
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 6
LENGTH: 236
TYPE: PRT
ORGANISM: *Tamias sibiricus*
US-09-140-804-6

Query Match 9.3%; Score 79; DB 3; Length 236;
Best Local Similarity 23.5%; Pred. No. 0.78;
Matches 24; Conservative 14; Mismatches 28; Indels 36; Gaps 4;
QY 30 GPPKAKWGMGMSFSLPQACVGAAPMOTGSPNRDGSQMDGALGPRGWTAPASCH--- 86
DB 73 GPPG-----TPGKPGPPGMNGFPGLPGPPPGMT-VNCHSKG 109
QY 87 ---LPLROSALHPSPS-----LIVESIGSVCTTHPRCP 118
DB 110 TSAFAVKANELPPAPDSQVIFKEALHDAQGHFDLATGVFTCP 151

RESULT 13
US-09-686-838B-6
Sequence 6, Application US/09686838B
Patent No. 6482612
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
TITLE OF INVENTION: Adipocyte-Specific Protein Homologs
FILE REFERENCE: 97-49D1
CURRENT APPLICATION NUMBER: US/09/686,838B
PRIOR FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: US 09/140,804
PRIOR FILING DATE: 1998-08-26
PRIOR APPLICATION NUMBER: US 60/056,983
PRIOR FILING DATE: 1997-08-26
NUMBER OF SEQ ID NOS: 50
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 6
LENGTH: 236
TYPE: PRT
ORGANISM: *Tamias sibiricus*
US-09-686-838B-6

Query Match 9.3%; Score 79; DB 4; Length 236;
Best Local Similarity 23.5%; Pred. No. 0.78;
Matches 24; Conservative 14; Mismatches 28; Indels 36; Gaps 4;
QY 30 GPPKAKWGMGMSFSLPQACVGAAPMOTGSPNRDGSQMDGALGPRGWTAPASCH--- 86
DB 73 GPPG-----TPGKPGPPGMNGFPGLPGPPPGMT-VNCHSKG 109
QY 87 ---LPLROSALHPSPS-----LIVESIGSVCTTHPRCP 118
DB 110 TSAFAVKANELPPAPDSQVIFKEALHDAQGHFDLATGVFTCP 151

RESULT 14
US-09-252-991A-27852

Sequence 27852, Application US/09252591A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196,136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 27852
LENGTH: 272
TYPE: PRT
ORGANISM: *Pseudomonas aeruginosa*
US-09-252-991A-27852

Query Match 9.3%; Score 79; DB 4; Length 272;
Best Local Similarity 26.9%; Pred. No. 0.94;
Matches 28; Conservative 11; Mismatches 47; Indels 18; Gaps 5;
QY 60 TSPNRDGSQMDGALGPRGWTAPASCHLPU-----ROSALHPSPS-----LIVESIG-- 106
DB 61 SSSRSASNSNSGAPFSTWTPPTSSSAPAPACRRNCRHPSPSPNPAFRSIDQ 120
QY 107 SVCVTHPRCPSPNTOHYMLPNSLSQIRTIIFINGETSP 150
DB 121 TYCESPRPRRPGTADHRW-----RMAQPARLRRAPAPADP 159

RESULT 15
US-09-513-999C-5144
Sequence 5144, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
PATENT NO. 6783961
FILE REFERENCE: 59,US2,REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 5144
LENGTH: 75
TYPE: PRT
ORGANISM: *Homo sapiens*
FEATURE:
NAME/KEY: UNSURE
LOCATION: 68
OTHER INFORMATION: Xaa= * or Gly
US-09-513-999C-5144

Query Match 9.2%; Score 78; DB 4; Length 75;
Best Local Similarity 34.0%; Pred. No. 0.23;
Matches 17; Conservative 7; Mismatches 18; Indels 8; Gaps 3;
QY 40 GWMGMSFSLPQACVGAAPMOT--GSPNRDGSQMDGALGPRGWTAP--ASC 85
DB 29 GWMGRIQSOPCSPPGSPPLQNPORSVHKGKRGQP-----PKCWXTSYAAC 74

Search completed: October 24, 2004, 16:04:22
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: October 24, 2004, 16:00:27 ; Search time 70 Seconds

(without alignments)
703.019 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848
1 MFLHYIDGVFPIQLVREKLH.....LSQIRPTISIFINGETSPSH 152

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1364641 seqs, 323758627 residues

Total number of hits satisfying chosen parameters: 1364641

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubppa/PC07_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubppa/PC07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep:*
- 17: /cgn2_6/ptodata/2/pubppa/US11_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubppa/US11_NEW_PUB.pep:*
- 19: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep:*
- 20: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score distribution.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	846	99.8	152 10 US-09-977-418-8	Sequence 8, Appl1
2	846	99.8	152 10 US-09-977-033A-8	Sequence 8, Appl1
3	846	99.8	152 10 US-09-977-751C-8	Sequence 8, Appl1
4	846	99.8	152 10 US-09-977-639A-8	Sequence 8, Appl1
5	846	99.8	152 11 US-09-977-819B-8	Sequence 8, Appl1
6	312	36.8	101 10 US-09-977-418-76	Sequence 76, Appl1
7	312	36.8	101 10 US-09-977-033A-76	Sequence 76, Appl1
8	312	36.8	101 10 US-09-977-751C-76	Sequence 76, Appl1
9	312	36.8	101 10 US-09-977-639A-76	Sequence 76, Appl1
10	312	36.8	101 11 US-09-977-819B-76	Sequence 76, Appl1
11	95	11.2	247 14 US-10-104-047-3142	Sequence 1142, Ap
12	92	10.8	210 16 US-10-437-963-190506	Sequence 190506,
13	90.5	10.7	502 14 US-10-029-386-32930	Sequence 32930, A

14	88.5	10.4	92	16	US-10-437-963-113445	Sequence 113445,
15	88.5	10.4	214	16	US-10-437-963-137051	Sequence 137051,
16	88	10.4	1745	10	US-09-795-061-4	Sequence 4, Appl1
17	86	10.1	306	14	US-10-156-761-7750	Sequence 7750, Ap
18	85	10.0	1106	14	US-10-369-493-5709	Sequence 5709, Ap
19	84.5	10.0	232	15	US-10-264-049-2379	Sequence 2379, Ap
20	84	9.9	138	15	US-10-424-599-183864	Sequence 183864,
21	83.5	9.8	631	15	US-10-433-544-12	Sequence 12, Appl1
22	82.5	9.7	68	14	US-10-106-598-5282	Sequence 5282, Ap
23	81.5	9.6	1279	13	US-10-087-192-1455	Sequence 1455, Ap
24	81	9.6	1739	10	US-09-795-061-2	Sequence 2, Appl1
25	80.5	9.5	1554	15	US-10-276-774-2260	Sequence 2260, Ap
26	80.5	9.5	3298	14	US-10-149-819-21	Sequence 21, Appl1
27	80.5	9.5	3312	14	US-10-225-567A-656	Sequence 656, App
28	80.5	9.5	3312	15	US-10-038-854-67	Sequence 67, Appl1
29	80.5	9.5	3312	15	US-10-038-854-67	Sequence 1399, Ap
30	80.5	9.5	4115	15	US-10-408-765A-1499	Sequence 4, Appl1
31	80	9.4	361	14	US-10-038-854-4	Sequence 68, Appl1
32	80	9.4	365	15	US-10-351-334-322	Sequence 322, App
33	80	9.4	542	14	US-10-411-120-86	Sequence 96, Appl1
34	80	9.4	542	14	US-10-351-334-323	Sequence 333, App
35	80	9.4	583	16	US-10-408-765A-1900	Sequence 1900, Ap
36	80	9.4	920	14	US-10-236-055A-26	Sequence 26, Appl1
37	80	9.4	1016	16	US-10-408-765A-1957	Sequence 1957, Ap
38	79.5	9.4	132	16	US-10-437-963-159554	Sequence 15954,
39	79	9.3	215	9	US-09-911-176B-51	Sequence 51, Appl1
40	79	9.3	215	14	US-10-160-762-54	Sequence 54, Appl1
41	79	9.3	215	14	US-10-222-857-1	Sequence 1, Appl1
42	79	9.3	215	14	US-10-360-186-54	Sequence 54, Appl1
43	79	9.3	236	14	US-10-197-293-6	Sequence 6, Appl1
44	78.5	9.3	268	14	US-10-185-425-6	Sequence 6, Appl1
45	78.5	9.3	402	14	US-10-343-953-2	Sequence 2, Appl1

ALIGNMENTS

US-09-977-418-8

RESULT 1

Sequence 8, Application US/09977418

Publication No. US20030027158A1

GENERAL INFORMATION:

APPLICANT: Shimkets et al

TITLE OF INVENTION: No. US20030027158A1 polynucleotides and polypeptides encoded t

FILE REFERENCE: 15966-552

CURRENT APPLICATION NUMBER: US/09/977,418

CURRENT FILING DATE: 2001-10-15

PRIOR APPLICATION NUMBER: 09/584,411

PRIOR FILING DATE: 2000-05-31

PRIOR APPLICATION NUMBER: USSN 60/189,810

PRIOR FILING DATE: 2000-03-16

PRIOR APPLICATION NUMBER: USSN 60/191,158

PRIOR FILING DATE: 2000-03-22

PRIOR APPLICATION NUMBER: USSN 60/193,086

PRIOR FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: USSN 60/201,398

PRIOR FILING DATE: 2000-05-03

NUMBER OF SEQ ID NOS: 93

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 8

LENGTH: 152

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: Variant

LOCATION: (1)-(152)

OTHER INFORMATION: where Xaa can any amino acid

US-09-977-418-8

Query Match 99.8%; Score 846; DB 10; Length 152;

Best Local Similarity 100.0%; Pred. No. 1e-73;

Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPHLYIDGVFPPIQLVREKLTETPAVLRLGPPGKAKMGWGSFSLPQACVYGAAPMOT 60
DB 1 MPHLYIDGVFPPIQLVREKLTETPAVLRLGPPGKAKMGWGSFSLPQACVYGAAPMOT 60
QY 61 GSPNRDSQGMGALGPRGWTTPASCHLPLRQSAALHPSPLIYESISVCVTHPRCPSPN 120
DB 61 GSPNRDSQGMGALGPRGWTTPASCHLPLRQSAALHPSPLIYESISVCVTHPRCPSPN 120
QY 121 TOHYHMLRPNSELSQIRFTISIFINGETSPSH 152
DB 121 TOHYHMLRPNSELSQIRFTISIFINGETSPSH 152

RESULT 2
US-09-977-033A-8
Sequence 8, Application US/09977033A
Publication No. US20030082554A1
GENERAL INFORMATION:
APPLICANT: Shimkels, Richard A
APPLICANT: Fernandes, Elma
APPLICANT: Herrman, John
APPLICANT: Vermet, Corine
TITLE OF INVENTION: No. US20030082554A1el nucleic acid sequences encoding human KIAA
TITLE OF INVENTION: protein-like and human protein PROZ28-like
FILE REFERENCE: 15966-552 CON-824
CURRENT APPLICATION NUMBER: US/09/977,033A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 152
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: variation
LOCATION: (53)
OTHER INFORMATION: wherein Xaa is any amino acid as defined by the specification
US-09-977-033A-8

Query Match 99.8%; Score 846; DB 10; Length 152;
Best Local Similarity 100.0%; Pred. No. 1e-73;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPHLYIDGVFPPIQLVREKLTETPAVLRLGPPGKAKMGWGSFSLPQACVYGAAPMOT 60
DB 1 MPHLYIDGVFPPIQLVREKLTETPAVLRLGPPGKAKMGWGSFSLPQACVYGAAPMOT 60
QY 61 GSPNRDSQGMGALGPRGWTTPASCHLPLRQSAALHPSPLIYESISVCVTHPRCPSPN 120
DB 61 GSPNRDSQGMGALGPRGWTTPASCHLPLRQSAALHPSPLIYESISVCVTHPRCPSPN 120
QY 121 TOHYHMLRPNSELSQIRFTISIFINGETSPSH 152
DB 121 TOHYHMLRPNSELSQIRFTISIFINGETSPSH 152

RESULT 3
US-09-977-751C-8
Sequence 8, Application US/09977751C
Publication No. US20030134430A1

GENERAL INFORMATION:
APPLICANT: Shimkels, Richard A
APPLICANT: Fernandes, Elma
APPLICANT: Herrman, John
APPLICANT: Vermet, Corine
TITLE OF INVENTION: No. US20030134430A1el Amino Acid Sequences for Human Caenorhabdit
TITLE OF INVENTION: Polypeptides.
FILE REFERENCE: 15966-552 CON 8-40
CURRENT APPLICATION NUMBER: US/09/977,751C
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 152
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Variant
LOCATION: (53)
OTHER INFORMATION: wherein Xaa is any amino acid as defined in the specification.
US-09-977-751C-8

Query Match 99.8%; Score 846; DB 10; Length 152;
Best Local Similarity 100.0%; Pred. No. 1e-73;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPHLYIDGVFPPIQLVREKLTETPAVLRLGPPGKAKMGWGSFSLPQACVYGAAPMOT 60
DB 1 MPHLYIDGVFPPIQLVREKLTETPAVLRLGPPGKAKMGWGSFSLPQACVYGAAPMOT 60
QY 61 GSPNRDSQGMGALGPRGWTTPASCHLPLRQSAALHPSPLIYESISVCVTHPRCPSPN 120
DB 61 GSPNRDSQGMGALGPRGWTTPASCHLPLRQSAALHPSPLIYESISVCVTHPRCPSPN 120
QY 121 TOHYHMLRPNSELSQIRFTISIFINGETSPSH 152
DB 121 TOHYHMLRPNSELSQIRFTISIFINGETSPSH 152

RESULT 4
US-09-977-639A-8
Sequence 8, Application US/09977639A
Publication No. US20030199103A1el amino acid sequences for human epidermal gr
GENERAL INFORMATION:
APPLICANT: Shimkels, Richard A
APPLICANT: Fernandes, Elma
APPLICANT: Herrman, John
APPLICANT: Vermet, Corine
TITLE OF INVENTION: No. US20030199103A1el amino acid sequences for human epidermal gr
TITLE OF INVENTION: polypeptides.
FILE REFERENCE: 15966-552 CON-834
CURRENT APPLICATION NUMBER: US/09/977,639A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03/584,411
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 8
LENGTH: 152
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Variant
LOCATION: (53)...(53)
OTHER INFORMATION: wherein Xaa is any amino acid as defined in the specification.
US-09-977-639A-8

Query Match 99.8%; Score 846; DB 10; Length 152;
Best Local Similarity 100.0%; Pred. No. 1e-73;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPEHYIDGVFPPIQVREKLTETPAVRLIGPPGKAKMGWGSFSLPQACVXGAAPMOT 60
DB 1 MPEHYIDGVFPPIQVREKLTETPAVRLIGPPGKAKMGWGSFSLPQACVXGAAPMOT 60
QY 61 GSPNRDGSQGDGALGPRGWTASCHLPRLRSALHPSPSLYESIGSVCTTHPRCPSPN 120
DB 61 GSPNRDGSQGDGALGPRGWTASCHLPRLRSALHPSPSLYESIGSVCTTHPRCPSPN 120
QY 121 TQHYHMLRPNSELSQIRFTISIFINGETSPSH 152
DB 121 TQHYHMLRPNSELSQIRFTISIFINGETSPSH 152

RESULT 5
US-09-977-819B-8
Sequence 8, Application US/09977819B
Publication No. US20040002134A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Fernandes, Elma
APPLICANT: Herrman, John
TITLE OF INVENTION: Corine
TITLE OF INVENTION: No. US20040002134A1el nucleic acid sequences encoding human KIAA
TITLE OF INVENTION: protein-like and human protein PRO228-like
FILE REFERENCE: 15966-552 CON-826
CURRENT APPLICATION NUMBER: US/09/977,819B
PRIOR FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 8
LENGTH: 152
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (53)...(53)
OTHER INFORMATION: "Xaa" = "Ile", "Leu", "Val" or "Phe"
US-09-977-819B-8
Query Match 99.8%; Score 846; DB 11; Length 152;

Best Local Similarity 100.0%; Pred. No. 1e-73;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPEHYIDGVFPPIQVREKLTETPAVRLIGPPGKAKMGWGSFSLPQACVXGAAPMOT 60
DB 1 MPEHYIDGVFPPIQVREKLTETPAVRLIGPPGKAKMGWGSFSLPQACVXGAAPMOT 60
QY 61 GSPNRDGSQGDGALGPRGWTASCHLPRLRSALHPSPSLYESIGSVCTTHPRCPSPN 120
DB 61 GSPNRDGSQGDGALGPRGWTASCHLPRLRSALHPSPSLYESIGSVCTTHPRCPSPN 120
QY 121 TQHYHMLRPNSELSQIRFTISIFINGETSPSH 152
DB 121 TQHYHMLRPNSELSQIRFTISIFINGETSPSH 152

RESULT 6
US-09-977-418-76
Sequence 76, Application US/09977418
Publication No. US20030027158A1
GENERAL INFORMATION:
APPLICANT: Shimkets et al
TITLE OF INVENTION: No. US20030027158A1el polynucleotides and polypeptides encoded t
FILE REFERENCE: 15966-552
CURRENT APPLICATION NUMBER: US/09/977,418
PRIOR FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: USSN 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: USSN 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: USSN 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: USSN 60/201,388
PRIOR FILING DATE: 2000-05-03
NUMBER OF SEQ ID NOS: 93
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 76
LENGTH: 101
TYPE: PRT
ORGANISM: Homo sapiens
US-09-977-418-76

Query Match 36.8%; Score 312; DB 10; Length 101;
Best Local Similarity 60.0%; Pred. No. 2.1e-22;
Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

QY 1 MPEHYIDGVFPPIQVREKLTETPAVRLIGPPGKAKMGWGSFSLPQACVXGAAPMOT 60
DB 1 MPEHYIDGVFPPIQVREKLTETPAVRLIGPPGKAKMGWGSFSLPQACVXGAAPMOT 60
QY 61 GSPNRDGSQGDG---ALGPRGWTASCHLPRLRSALHPSPSLI 101
DB 61 DHLEWEARAWMGLWLEVGPO---LLATPSGSGALHPSPSLI 101

RESULT 7
US-09-977-033A-76
Sequence 76, Application US/09977033A
Publication No. US20030082554A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Fernandes, Elma
APPLICANT: Herrman, John
TITLE OF INVENTION: Corine
TITLE OF INVENTION: No. US20030082554A1el nucleic acid sequences encoding human KIAA
TITLE OF INVENTION: protein-like and human protein PRO228-like
FILE REFERENCE: 15966-552 CON-824
CURRENT APPLICATION NUMBER: US/09/977,033A
PRIOR FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322

PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 76
LENGTH: 101
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-033A-76

Query Match 36.8%; Score 312; DB 10; Length 101;
Best Local Similarity 60.0%; Pred. No. 2.1e-22;
Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

CY 1 MPEHYIDGVPPIDLVREKLTETPAVLRLLGPPGKAKGKMGWGSFSLPQACVXGAAPMOT 60
DB 1 MPEHYIDGVPPIDLVREKLTETPAVLRLLGPPGKAKGKMGWGSFSLPQACVXGAAPMOT 60
CY 61 GSPNRDSOGMDG---ALGPRGWTASCHLPQRSALHPSPLI 101
DB 61 DILTEMEARAMGMLWLVGVGPQ---LLATFPGSGSLHPSPLI 101

RESULT 8
US-09-977-751C-76

Sequence 76, Application US/09977751C
Publication No. US20030134430A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Herrman, John
APPLICANT: Vernet, Corine
TITLE OF INVENTION: No. US20030134430A1et Amino Acid Sequences for Human Caenothabdit
FILE REFERENCE: 15966-552 CON S-40
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 76
LENGTH: 101
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-751C-76

Query Match 36.8%; Score 312; DB 10; Length 101;
Best Local Similarity 60.0%; Pred. No. 2.1e-22;
Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

CY 1 MPEHYIDGVPPIDLVREKLTETPAVLRLLGPPGKAKGKMGWGSFSLPQACVXGAAPMOT 60
DB 1 MPEHYIDGVPPIDLVREKLTETPAVLRLLGPPGKAKGKMGWGSFSLPQACVXGAAPMOT 60
CY 61 GSPNRDSOGMDG---ALGPRGWTASCHLPQRSALHPSPLI 101
DB 61 DILTEMEARAMGMLWLVGVGPQ---LLATFPGSGSLHPSPLI 101

RESULT 9
US-09-977-639A-76

Sequence 76, Application US/09977639A
Publication No. US20030199103A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Herrman, John
APPLICANT: Vernet, Corine
TITLE OF INVENTION: No. US20030199103A1et amino acid sequences for human epidermal gr
FILE REFERENCE: 15966-552 CON S-34
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 76
LENGTH: 101
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-639A-76

Query Match 36.8%; Score 312; DB 10; Length 101;
Best Local Similarity 60.0%; Pred. No. 2.1e-22;
Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

CY 1 MPEHYIDGVPPIDLVREKLTETPAVLRLLGPPGKAKGKMGWGSFSLPQACVXGAAPMOT 60
DB 1 MPEHYIDGVPPIDLVREKLTETPAVLRLLGPPGKAKGKMGWGSFSLPQACVXGAAPMOT 60
CY 61 GSPNRDSOGMDG---ALGPRGWTASCHLPQRSALHPSPLI 101
DB 61 DILTEMEARAMGMLWLVGVGPQ---LLATFPGSGSLHPSPLI 101

RESULT 10

US-09-977-819B-76
Sequence 76, Application US/09977819B
Publication No. US20040002134A1
GENERAL INFORMATION:
APPLICANT: Shimkets, Richard A
APPLICANT: Herrman, John
APPLICANT: Vernet, Corine
TITLE OF INVENTION: No. US20040002134A1et nucleic acid sequences encoding human KIAA0
TITLE OF INVENTION: protein-like and human protein PRO228-like
FILE REFERENCE: 15966-552 CON S-26
CURRENT APPLICATION NUMBER: US/09/977,819B

```

; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 60/137,322
; PRIOR FILING DATE: 1999-06-03
; PRIOR APPLICATION NUMBER: 60/189,810
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/191,158
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/193,086
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/201,388
; PRIOR FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: 09/584,411
; PRIOR FILING DATE: 2000-05-31
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 76
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-819B-76

```

```

Query Match          36.8%; Score 312; DB 11; Length 101;
Best Local Similarity 60.0%; Pred. No. 2,1e-22;
Matches 63; Conservative 7; Mismatches 27; Indels 8; Gaps 2;

```

```

CY 1 MPELYIDGVPIQLVREKHEHTETPAVRLILGPPGAKMGWGSFSLPQACVXGAAPMOT 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 MPELYIDGVPIQLVREKHEHTETPAVRLILGPPGAKMGWGSFSLPQACVXGAAPMOT 60
CY 61 GSPNRDGSQMDG---ALGPRGWTTPASCHLILNQSALHPSPSLI 101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 DHEEMERAMMGLWLEVGFPQ---LLATPSSQSALHPSPSLI 101

```

```

RESULT 11
US-10-104-047-3142
; Sequence 3142; Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3142
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3142

```

```

Query Match          11.2%; Score 95; DB 14; Length 247;
Best Local Similarity 25.9%; Pred. No. 0.49;
Matches 35; Conservative 11; Mismatches 23; Indels 60; Gaps 7;

```

```

CY 2 PHLYIDGVPIQLVREKHEHTETPAVRLILGPPGAKMGWGSFSLPQACVXGAAPMOT 37
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 105 PHRVVAVLVAARQLTLEQLHAVEISAVHSRDHGLEVGPPGPRPPRGIALAFWFAPMGNT 164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CY 38 -----GM-----GMGWSFSLPQACVXGAAPMOTGSPNR 65
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 165 SPEGSLAGLGRVSLSPAMGWGEGSGAGRPKKGKGMGSPALPLPGCTVTSALLPAINP-- 222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CY 66 DSGQMDGALGPRGW 80
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 223 -GSGGV-GSVGRKGM 235

```

```

RESULT 12
US-10-437-963-190506
; Sequence 190506; Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 190506
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_86914C.1.pap
US-10-437-963-190506

```

```

Query Match          10.8%; Score 92; DB 16; Length 210;
Best Local Similarity 26.2%; Pred. No. 0.79;
Matches 39; Conservative 14; Mismatches 50; Indels 46; Gaps 6;

```

```

CY 21 ETPAVRLILGPPGAKMGWGSFSLPQACVXGAAPMOTGSPNRDGSQMDGALG--- 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14 ESRSPAMMGPPGSGRGMGVEIARR--WGRITVSGTARVGRGMIGRLGYVAS 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CY 77 ---PRGWTPASCHLILNQSALHPSPSLIYE-----SIGVCVTT 112
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 72 VSCLSGMCAGSSWVP-RSSAYPSVSPSSSAYDLPTGTLTGNGSSASVPHSLGLHCVL- 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
CY 113 HPRCPSPNTOHMYLAPNSELGIRFTIS 141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 130 -----LTGNAEGHAFWTV 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
US-10-029-386-32930
; Sequence 32930; Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32930
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC012313.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: P51523, EVALUATE 0.00e+00
US-10-029-386-32930

```

Query Match 10.7%; Score 90.5; DB 14; Length 502;
Best Local Similarity 31.1%; Pred. No. 2.9;
Matches 42; Conservative 7; Mismatches 39; Indels 47; Gaps 11;

QY 33 GKAKWNG-WGMSFSLP-----QACVGAAPMGTGSPNRD---GSGMGALGPRGW 80
DB 36 GAVSGGGAWEKSTEVREXGQOQATLGADEQ-GGGRGLGADGGRDGA-GPRS- 92
QY 81 TPASCHLPLRQSLHPSFSLIYESIGSVG-----VTTT-----PYRCPEN 120
DB 93 EPA-----DRALRPSP--LPEEPGRCGCGKARSGSYLLQHRVATGKPYTCPEC 143
QY 121 TQHYMLRPNSELSQ 135
DB 144 GKAFW---SSNLSQ 155

RESULT 14
US-10-437-963-113445
Sequence 113445, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437, 963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 113445
LENGTH: 92
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_17231C.1.pap
US-10-437-963-113445

Query Match 10.4%; Score 88.5; DB 16; Length 92;
Best Local Similarity 29.7%; Pred. No. 0.69;
Matches 27; Conservative 14; Mismatches 15; Indels 35; Gaps 7;

QY 30 GPPGKAKWNG-----WGMSFSLPQAC-----VXGAAPMGTGSPNRDGSQ 69
DB 2 GPP-----WGWGHSPPRPRGWSGMNPNRQCRVSHRGRRGTGAC-VRSIDPSIDRAR 56
QY 70 GMDGALGPRGWTPLRQSLHPSFSLIYESIGSVG-----VTTT-----PYRCPEN 120
DB 57 G--GVISP---TPSA-----RSSAGHRSPAT 77

RESULT 15
US-10-437-963-137051
Sequence 137051, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B

CURRENT APPLICATION NUMBER: US/10/437, 963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 137051
LENGTH: 214
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_38571C.1.pap
US-10-437-963-137051

Query Match 10.4%; Score 88.5; DB 16; Length 214;
Best Local Similarity 25.6%; Pred. No. 1.8;
Matches 40; Conservative 17; Mismatches 70; Indels 29; Gaps 7;

QY 3 HLYIDG-VFPIQLVREKLETPAVLRLLGPPGKAKWNGWSFSLPQACVXGAAPMGTG 61
DB 45 HLIVSGSSFPVARTELGHARPTWPLLRPPSSRMTW-----PLQASGGGKETE 96
QY 62 SPNDGSGMDGALGPRGWTPLA-----SCHL-PLRQSLHPSFSLIYESIGSVCTT 113
DB 97 EGCGCGGGGHSCEP--TPADMFLVSCRRPPLRTVSTAVALALALAT-TFLTSH 152
QY 114 PYRCPNTQHYMLRPNSELSQIRFTISIFINGETS 149
DB 153 ARSSAH-----PRAQLPTFAFASATLIGCTS 180

Search completed: October 24, 2004, 16:05:37
Job time : 71 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 24, 2004, 15:41:11 ; Search time 24 Seconds

(without alignments)
609.373 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848
Sequence: 1 MPHLYIDGVFPPIQLVREKLM.....LSQIRFTISIFINGETSPSH 152

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
PIR 79: *
1: P1r1.*
2: P1r2.*
3: P1r3.*
4: P1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93.5	11.0	650	2	T22002
2	85	10.0	1106	2	T18739
3	83	9.8	219	2	A35650
4	82	9.7	445	2	S43492
5	81	9.6	423	2	T16750
6	80.5	9.5	1364	2	T00250
7	80	9.4	992	2	T08772
8	79	9.3	215	2	C48150
9	77.5	9.1	330	2	A40855
10	77.5	9.1	544	2	S62527
11	77.5	9.1	555	2	D95377
12	77.5	9.1	584	2	I50419
13	77	9.1	68	2	S47648
14	77	9.1	290	2	B88638
15	77	9.1	404	1	I51596
16	77	9.1	1120	2	H88449
17	76	9.1	1690	1	CGH018
18	76	9.0	1419	1	A41182
19	76	9.0	1486	1	B40333
20	76	9.0	1487	2	B41182
21	75.5	8.9	325	2	S02170
22	75.5	8.9	337	2	T23794
23	75.5	8.9	498	2	A48203
24	75.5	8.9	1744	2	S40991
25	75	8.8	460	2	T33110
26	74.5	8.8	1603	2	S23810
27	74.5	8.8	107	2	B61396
28	74.5	8.8	342	2	A46396
29	74.5	8.8	921	2	S40495

ALIGNMENTS

30	74.5	8.8	921	2	S42617	collagen alpha 1(I
31	74	8.7	313	2	T33010	hypothetical prote
32	74	8.7	1187	2	TVHUY2	proteih-tyrosine k
33	74	8.7	1747	2	A45974	collagen alpha 1(X
34	74	8.7	1857	2	S31212	collagen alpha 1(X
35	74	8.7	1888	2	S78476	collagen alpha 1(X
36	73.5	8.7	509	2	H87389	conserved hypothet
37	73.5	8.7	931	2	S13580	collagen alpha 1(I
38	73.5	8.7	1132	2	T03844	telomerase catalyt
39	73.5	8.7	1546	1	CGH02E	collagen alpha 2(X
40	73.5	8.7	1570	1	CGH03B	collagen alpha 3(I
41	73.5	8.7	1591	1	S22917	collagen alpha 5(I
42	73	8.6	121	2	F72580	hypothetical prote
43	73	8.6	198	2	I49558	collagen alpha 1(I
44	73	8.6	360	2	T37285	collagen dpy-2 - C
45	73	8.6	380	2	T28868	cuticle collagen d

RESULT 1

T22002 hypothetical protein F39H1.4 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C/Accession: T22002

R/White, S.

submitted to the EMBL Data Library, October 1996

A/Accession: T22002

A/Reference number: Z19500

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-650 <MIL>

A/Cross-references: UNIPROT:O17866; EMBL:Z81079; PIDN:CA803084.1; GSPDB:GN00019; CESP:F

A/Experimental source: clone F39H1.1

C/Genetics:

A/Map position: 1

A/Map position: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

A/Intons: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match

Best Local Similarity 38.6%; Pred. No. 0.42;

Matches 22; Conservative 6; Mismatches 24; Indels 5; Gaps 1;

DB 23 PAVRLRPPGKAKMGWGFSLPQACVXGAAPKQTSFPRDSCGMDALGPRG 79

229 PGMPGLAGPPPGSGCTGHSAGADGPQ-----GPGGLPGAPGRDGTSGVGGRGPGG 280

RESULT 2

T18739 hypothetical protein B0393.5 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C/Accession: T18739

R/Sultson, J.

submitted to the EMBL Data Library, September 1994

A/Accession: T18739

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1106 <MIL>

A/Cross-references: UNIPROT:Q17494; EMBL:Z37983; NID:el519039; PIDN:CA86058.1; GSPDB:G

A/Experimental source: clone B0393

C/Genetics:

A/Map position: 3

A/Map position: 15/3; 60/1; 100/3; 343/2; 443/1; 507/3; 609/3; 659/1; 796/1; 876/1; 946/1; 9

Query Match

Best Local Similarity 23.7%; Pred. No. 4.9;

Matches 36; Conservative 17; Mismatches 41; Indels 58; Gaps 7;

QY 12 IQLVREKLN-----ETPAVLRLI-----GPPGKAKWGMGMSFSLPQACV 52
 Db 722 LSLIVKVCQARGPMTGISTFSRFLMLPNRLASGSNVQCPGK-----MMYRVLDIDL 773
 QY 53 XGAAPMOTGSP-----NRDQSGMD-----GALGPPGMPASCH 86
 Db 774 QCTPPGGLGFLCDRCQAGHVCINCESTCHCGSVACDVTGMCPGALCRAGMGSSCD 833
 QY 87 LPLRQSLHPSPLIYESIGSVCTTH-PYRC 117
 Db 834 QDIDECHEM-----SLVMCAVGSQCCTVTRGVCRC 861

RESULT 3

Sur protein - chicken
 C:Species: Gallus gallus (chicken)
 C>Date: 28-Sep-1990 #sequence_revision 28-Sep-1990 #ext_change 09-Jul-2004
 C/Accession: A35650
 R:Dora, T.; Wang, L.H.
 Mol. Cell. Biol. 10, 4068-4079, 1990
 A>Title: An alternative non-tyrosine protein kinase product of the c-src gene in chicken
 A:Reference number: A35650; PMID:90318371; PMID:2115117
 A:Accession: A35650
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1219 <DOR>
 A:Cross-references: UNIPROT:Q99370; GB:M57290; NID:G212703; PIND:AAA9076.1; PID:G212704

Query Match 9.8%; Score 83; DB 2; Length 219;
 Best Local Similarity 22.1%; Pred. No. 1.3;
 Matches 29; Conservative 15; Mismatches 39; Indels 46; Gaps 6;

QY 40 GWMGNSLPOACVXGAAPMOTGSP-----NRDQSGMD-----GALG-----PRG 79
 Db 10 GVGW-ERRPIASGSLQLHLGPMRRHGFVLGRDGMGRDVLGAFGMPSLGEAGVPGA 67
 QY 80 WTPASCHLPLR-----QSLHPSPLIYESIGSVCTTHPYR----- 116
 Db 68 AAVLELMIKGVLVLQCTPRCSRCHHPSSPVVSKHLSTPSTTCPCRGTHRCAGGQC 127
 QY 117 ---CPSTQHY 124
 Db 128 AVPCPAGSCH 138

RESULT 4

S43492
 Surface antigen - hepatitis B virus (subtype adr)
 N/Alternate names: HBs antigen
 N/Contains: surface antigen pre-S1 (large envelope protein); surface antigen pre-S2 (mid
 C/Species: hepatitis B virus, HBV
 A/Variety: subtype adr
 C/Date: 07-Sep-1994 #sequence_revision 10-Nov-1995 #ext_change 09-Jul-2004
 R/Accession: S43492; PNO601; PNO602
 R/Jongarevic, I.F.; Zengraf, H.; Schroeder, C.H.
 Nucleic Acids Res. 18, 4940, 1990
 A>Title: Sequence of a replication competent hepatitis B virus genome with a prex open r
 A:Reference number: S12598; PMID:90370503; PMID:2395664
 A:Accession: S43492
 A>Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1445 <LON>
 A/Cross-references: UNIPROT:Q67953; EMBL:X52939; NID:G457780; PIND:CAA37114.1; PID:G4577
 A/Experimental source: subtype adr
 R/Mime, L.T.; Solomon, L.R.; Ebert, J.W.; Fields, H.
 Biochem. Biophys. Res. Commun. 195, 186-191, 1993
 A>Title: Unique prex sequence in a gibbon-derived hepatitis B virus variant.
 A:Reference number: PNO601; PMID:93371402; PMID:863598
 A:Accession: PNO601
 A:Molecule type: protein
 A:Residues: 57-95 <MIM>

A/Accession: PNO602
 A:Molecule type: protein
 A:Residues: 165-196 <MI2>
 C/Genetics:
 A:Gene: S
 A:Introns: 210/3
 C/Superfamily: hepatitis B virus surface antigen
 C/Keywords: surface antigen
 F/1-445/Product: surface antigen pre-S1 (large envelope protein) #status predicted <PS1>
 F/1-164/Domain: pre-S1 domain #status predicted <PRE1>
 F/165-445/Product: surface antigen pre-S2 (middle envelope protein) #status predicted <P
 F/165-219/Domain: pre-S2 domain #status predicted <PRE2>
 F/220-445/Product: surface antigen S (small envelope protein) #status predicted <PSD>

Query Match 9.7%; Score 82; DB 2; Length 445;
 Best Local Similarity 25.0%; Pred. No. 3.5;
 Matches 40; Conservative 14; Mismatches 48; Indels 58; Gaps 9;

QY 15 VREKHEETPAVL---RLGPPGKAKWGW-----GWMGNSLPOAC-----VXGA- 55
 Db 23 IREKLAHPHVFVGHILGNKSYGMGMSKPRGQMTNLSVPRPLGFPDPHQLDPAFGAN 82
 QY 56 -APMOTGSPNRD---GSGQMDGALGP-----RGWT-----PASC 85
 Db 83 SNNEDMDFNPNKQHWPEAKQVGAFAFGPFTPHGHLGWSPOAGTLTTPAAPPPAST 142
 QY 86 HLPRLQSLHPSPLIYESIGSVCTTHPYRCPSNTQHY 125
 Db 143 N---RQSGROPTP-----ISPLRDSHPAMQMNSTTH 173

RESULT 5

hypothetical protein R13F6.9 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #ext_change 15-Sep-2000
 C/Accession: T16750
 R/Miller, N.
 Submitted to the EMBL Data Library, April 1994
 A>Description: The sequence of C. elegans cosmid R13F6.
 A:Reference number: Z18570
 A:Accession: T16750
 A>Status: preliminary; translated from GB/EMBL/DDBT
 A:Molecule type: DNA
 A:Residues: 1-423 <ML>
 A/Cross-references: EMBL:U00046; NID:G470358; PID:G470367; PIND:AA47050.1; GSPDB:GN0002
 A/Experimental source: strain Bristol N2; clone R13F6
 C/Genetics:
 A:Gene: CESP.R13F6.9
 A:Map position: 3
 A:Introns: 47/1; 77/3; 102/3; 132/2; 154/1; 182/3; 214/3; 234/2; 318/3; 345/3; 372/3; 39
 C/Superfamily: human transcription regulator MBD-4

Query Match 9.6%; Score 81; DB 2; Length 423;
 Best Local Similarity 27.0%; Pred. No. 4.1;
 Matches 31; Conservative 16; Mismatches 42; Indels 26; Gaps 5;

QY 42 GWSFSLPOACVXGAAPMOTGSPNRDQSGMDGALGPRGWTPLRQSLHPSPLI 101
 Db 138 GLNWSMP-----SPQPISSPVTIWQSS-----GSSIASC-----ASSPSPSYF 175
 QY 102 YESIGSVCTTH--PYRCPSNTQHYMKLRNSLSOI--RFTSIFINGETSPSH 152
 Db 176 SEDGGEVQVQRPPFPHPSMAQITTFELNSRVSVEFKLVNLSITVDGTTNPSN 230

RESULT 6

T00250
 MEGF2 protein - human (fragment)
 C/Species: Homo sapiens (man)
 C/Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #ext_change 21-Jul-2003
 C/Accession: T00250
 R/Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomic 51, 27-34, 1998
 A>Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
 A:Reference number: Z14126; MUID:98360089; PMID:9693030
 A:Accession: T00250
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1364 <NAK>
 A:Cross-references: EMBL:AB011536; NID:G3449297; PIDN:BA32464.1; PID:G3449298
 A:Experimental source: brain; clone HG1044
 C:Genetics:
 A:Gene: MEGP2
 A:Map position: 3p21.2-p24.1
 F:1-28/Domain: EGF homology (fragment) <EGF>
 F:32-66/Domain: EGF homology <EGF1>
 F:124-169/Domain: laminin-type EGF-like homology <LEG>

Query Match 9.5%; Score 80.5; DB 2; Length 1364;
 Best Local Similarity 32.1%; Pred. No. 17;
 Matches 34; Conservative 4; Mismatches 33; Indels 35; Gaps 6;

QY 13 QLVREKLHETPA-VLRLLGPPGKAKMGWGSFSLPQACVXGA---APMOTGS-----62
 DB 1131 QLSERLEAPAPVLRPLSRGS-----QECMDAAPGRLEPKDGSSTLPR 1176

QY 63 -PNDGSGQNDGALG-----PRGWTPASCHLPKQSAHPS 98
 DB 1177 QPRDYFGAMAGRPSRDALDLAGRMTL-STLPPRRTRDLPQP 1221

RESULT 7
 T08772
 hypothetical protein DKFZ586M121.1 - human (fragment)
 C:Species: Homo sapiens (man)
 C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
 C:Accession: T08772
 R:Ottensmeyer, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, May 1999
 A:Reference number: Z16474
 A:Accession: T08772
 A:Molecule type: mRNA
 A:Residues: 1-992 <OTT>
 A:Cross-references: UNIPROT:Q9Y6C2; EMBL:AL050138
 A:Experimental source: adult uterus; clone DKFZ586M121
 C:Genetics:
 A>Note: DKFZ586M121.1

Query Match 9.4%; Score 80; DB 2; Length 992;
 Best Local Similarity 25.9%; Pred. No. 13;
 Matches 30; Conservative 12; Mismatches 30; Indels 44; Gaps 4;

QY 12 IQLVREKLHE-----TPAVLRLLGPPGKAKMGWGSFSLPQACVXGAAP 57
 DB 773 LQLIEDRHLQSLKDLGPPAGEAPPPGPGGPPGAG-----811

QY 58 MOTGSPNDSQGDGALGPRG-----WTPASCHLPKQSAHPSLIYESIGSV 108
 DB 812 -PPSPGKDGEGPIGPPGQEGVGAPAA---PVPQVAFSAALSPRSEPGTV 863

RESULT 8
 C48150
 hibernation-related protein HP-27 precursor - Siberian chipmunk
 C:Species: Eutamias sibiricus (Siberian chipmunk)
 C>Date: 16-Feb-1994 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 C:Accession: C48150; A41752
 R:Takamatsu, N.; Ohba, K.; Kondo, J.; Kondo, N.; Shiba, T.
 Mol. Cell. Biol. 13, 1516-1521, 1993
 A>Title: Hibernation-associated gene regulation of plasma proteins with a collagen-like
 A:Reference number: A48150; MUID:93180798; PMID:8441393
 A:Accession: C48150
 A:Molecule type: mRNA; protein
 A:Residues: 1-215 <TAK>
 A:Cross-references: UNIPROT:Q06577; GB:DI2976; NID:G287471; PIDN:BA02353.1; PID:G287472

A>Note: the source is designated as Tamias asiaticus in Genbank entry TWSHP27, release 1
 A>Note: sequence extracted from NCBI backbone (NCBIN:125948, NCBI:125949)
 R:Kondo, N.; Kondo, J.
 J. Biol. Chem. 267, 473-478, 1992
 A>Title: Identification of novel blood proteins specific for mammalian hibernation.
 A:Reference number: A41752; MUID:92112696; PMID:1730610
 A:Accession: A41752
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 31-40, 'A', 42-50, 'Q', 52-215 <KON>
 C:Superfamily: complement subcomponent C1q chain A; complement C1q carboxyl-terminal ho
 C:Keywords: glycoprotein; hibernation; plasma
 F:1-30/Domain: signal sequence status predicted <SIG>
 F:31-215/Product: hibernation-related protein HP-27 #status experimental <MAT>
 F:91-214/Domain: complement C1q carboxyl-terminal homology <C1Q>
 F:155/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.3%; Score 79; DB 2; Length 215;
 Best Local Similarity 23.5%; Pred. No. 3;
 Matches 24; Conservative 14; Mismatches 28; Indels 36; Gaps 4;

QY 30 GPFGKAKMGWGSFSLPQACVXGAAPWQTSPPNRDSQGDGALGPRGTPASCH---86
 DB 52 GPFG-----TPGKPPPPGMNGPFGLPGRPPGWT-VNCHSNG 88

QY 87 ---LPLRQSAHPSRS-----LIYESIGSVCTTHPRCP 118
 DB 89 TSAFAVKANLPPASQPVIFKALHDAQGHFDLATGVFTCP 130

RESULT 9
 A40855
 homeotic protein Hox 11 - human
 N:Alternate names: tcl-3 proto-oncogene
 C:Species: Homo sapiens (man)
 C>Date: 28-Feb-1992 #sequence_revision 28-Feb-1992 #text_change 09-Jul-2004
 C:Accession: A40855; S17311; A41224
 R:Hatanou, M.; Roberts, C.W.M.; Minden, M.; Crist, W.M.; Korsmeyer, S.J.
 Science 253, 79-82, 1991
 A>Title: Deregulation of a homeobox gene, HOX11, by the t(10;14) in T cell leukemia.
 A:Reference number: A40855; MUID:91289163; PMID:1676542
 A:Accession: A40855
 A:Molecule type: mRNA
 A:Residues: 1-330 <HAT>
 A:Cross-references: UNIPROT:P31314; GB:S38742; NID:G232582; PIDN:AA819293.1; PID:G23258
 R:Lu, M.; Gong, Z.; Shen, W.; Ho, A.D.
 EMBO J. 10, 2905-2910, 1991
 A>Title: The tcl-3 proto-oncogene altered by chromosomal translocation in T-cell leukem
 A:Reference number: S17311; MUID:92007734; PMID:1717256
 A:Accession: S17311
 A:Molecule type: mRNA
 A:Residues: 1-275, 'L', 277-330 <LUM>
 A:Cross-references: EMBL:M62626; NID:G339199; PIDN:AAA36719.1; PID:G387693
 R:Kennedy, K.A.; Gonzalez-Sarmiento, R.; Kees, U.R.; Lampert, F.; Dear, N.; Boehm, T.;
 Proc. Natl. Acad. Sci. U.S.A. 88, 8900-8904, 1991
 A>Title: HOX11, a homeobox-containing T-cell oncogene on human chromosome 10q24.
 A:Reference number: A41224; MUID:92020958; PMID:1681546
 A:Accession: A41224
 A:Molecule type: DNA
 A:Residues: 1-189, 'VALSPFTVTRRT', 190-275, 'L', 277-330 <KEN>
 A:Cross-references: GB:W75952
 A>Note: the authors suggest a long form with twelve additional amino acids from possibl
 C:Genetics:
 A:Gene: GDB:HOX11
 A:Cross-references: GDB:119607; OMIM:186770
 A:Map position: 10q24-10q24
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; leukemia; nucleus; proto-oncogene; transcription reg
 F:202-258/Domain: homeobox homology <HOX>

Query Match 9.1%; Score 77.5; DB 2; Length 330;
 Best Local Similarity 27.1%; Pred. No. 6;
 Matches 32; Conservative 10; Mismatches 59; Indels 17; Gaps 4;

QY 24 AYVLELPPGKAKWGMGWSFSL-----PQACVXGAAPMOTGSPNPD-----GSGQMGD 73
 Db 6 SVASPPGP--AAMGCMASGYAASRAQPPRHLPADPPHPPTRTQTGTGCHGGRGTG 63
 QY 74 ALGPR 78
 Db 64 RLGAR 68

RESULT 14
 B88638
 protein F58F6.2 [imported] - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
 C/Accession: B88638
 R/Anonymous: The C. elegans Sequencing Consortium.
 S/Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A/Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A/Reference number: A75000; MUID:99069613; PMID:9851916
 A/Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
 A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A/Accession: B88638
 A/Status: Preliminary
 A/Molecule type: DNA
 A/Residues: 1-290 <STO>
 A/Cross-references: UNIPROT:O44173; GB:chr_IV; PID:AA88359.1; PID:g2662600; GSPDB:GNO0
 C/Genetics:
 A/Note: Similar to cuticular collagen
 A/Gene: F58F6.2
 A/Map position: 4

Query Match 9.1%; Score 77; DB 2; Length 280;
 Best Local Similarity 37.7%; Pred. No. 6.4;
 Matches 23; Conservative 1; Mismatches 19; Indels 18; Gaps 3;

QY 30 GPPGKAKWGMGWSFSLPQACVXGAAPMOTGSPNPDGSGQMGALPPRGWT--PASCHL 87
 Db 142 GPPGKPG-----PQ-----GAPGDIGQPGNPGRMDEPDPGPPTGPPGPRGL 185
 QY 88 P 88
 Db 186 P 186

RESULT 15
 I61596
 advanced glycosylation end-products receptor precursor - human
 N/Alternate names: advanced glycosylation end product-binding protein, 35k; glycoprotein
 C/Species: Homo sapiens (man)
 C/Date: 24-May-1996 #sequence_revision 07-Feb-1997 #text_change 09-Jul-2004
 C/Accession: I61596; B42879; S27968
 R/Sugaya, K.; Fukagawa, T.; Matsumoto, K.; Mita, K.; Takahashi, E.; Ando, A.; Inoko, H.;
 Genomics 23, 408-419, 1994
 A/Title: Three genes in the human MHC class III region near the junction with the class
 II region of mouse mammary tumor gene int-3.
 A/Reference number: A55562; MUID:95137587; PMID:7835890
 A/Accession: I61596
 A/Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL
 A/Molecule type: DNA
 A/Residues: 1-404 <RES>
 A/Cross-references: UNIPROT:O15109; GB:D8769; NID:G561657; PID:BA05958.1; PID:G561659
 R/Neepert, M.; Schmidt, A.M.; Bretz, U.; Yan, S.D.; Wang, F.; Pan, Y.C.; Elliston, K.; St
 u. Biol. Chem. 267, 14996-15004, 1992
 A/Title: Cloning and expression of a cell surface receptor for advanced glycosylation en
 A/Reference number: A42879; MUID:92340547; PMID:1378843
 A/Accession: B42879
 A/Molecule type: mRNA
 A/Residues: 'G', '2-99', 'R', '101-404' <NEE>
 A/Cross-references: EMBL:M91211; NID:g190845; PIDN:AAA03574.1; PID:g190846
 A/Experimental source: lung
 A/Note: sequence extracted from NCBI backbone (NCEIP:109438)
 C/Comment: Advanced glycosylation end products are heterogeneous nonenzymatically glycos

cellular function, thus contributing to tissue lesions in diabetes.
 C/Comment: This receptor appears also to mediate the effects of amyloid beta peptide on
 ates in the neurotoxic pathway that produces dementia in Alzheimer's disease.

A/Genetics:
 A/Gene: GDB:AGER
 A/Cross-references: GDB:306354; OMIM:600214
 A/Map position: 6p21.3-6p21.3
 A/Introns: 18/1; 53/3; 119/1; 140/3; 170/1; 231/1; 274/3; 322/1; 331/1; 373/2
 C/Function:
 A/Description: neuronal receptor for amphoterin, a DNA-binding protein involved in neur
 C/Superfamily: advanced glycosylation end products receptor; immunoglobulin homology
 C/Keywords: Alzheimer's disease; glycoprotein; receptor; transmembrane protein
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-404/Product: advanced glycosylation end products receptor #status predicted <MAT>
 F:23-344/Domain: extracellular #status predicted <EXT>
 F:31-101/Domain: immunoglobulin homology <IM1>
 F:137-210/Domain: immunoglobulin homology <IM2>
 F:252-303/Domain: immunoglobulin homology <IM3>
 F:345-362/Domain: transmembrane #status predicted <TM>
 F:363-404/Domain: intracellular #status predicted <INT>
 F:25, 81/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:38-99, 144-208, 259-301/Disulfide bonds: #status predicted

Query Match 9.1%; Score 77; DB 1; Length 404;
 Best Local Similarity 23.3%; Pred. No. 9.3;
 Matches 41; Conservative 16; Mismatches 57; Indels 62; Gaps 7;

QY 8 GVPPIQVREKHEPPAVRLRLLGPPGKAKWGMGWSFSLPQACVXGAAPM----- 59
 Db 184 GLFTIQ---SELMVTPA-----RGDPRPTSCSFSPGLFPHRLRTPAPDPVWBPVP 234
 QY 60 -----TGSPPNRDGSQMGALGPPRGWTPASCHLPLRQS-----ALHPSPLI 101
 Db 235 LEEVQLVPEP-----GGAVALPGGTVLTLEVAQPSPLQIMKMDGVLPLPSPVLI 287
 QY 102 YESIGS-----VCTTHPRCPSTNQHMLRPSLSQIFTSIFNGSTSS 151
 Db 288 LPEIGPDQGTYSVATSHSHGPOESR-----AVSILIEPGSGEPT 329

Search completed: October 24, 2004, 16:03:52
 Job time : 26 secs

QY 74 A-----LGPR-GWTPAS-----CHLPLRQSLHSPS-- 98
 DB 272 ARHRELESVPLRGVKTATPKRTWPGPFTFABLRATCTGCHRSFEAKLPFCPOC 331
 QY 99 -SLIYESIGSVCTTHPYRCPSNTQYHW-----LRPNSLSQRTFTSIFINGET 148
 DB 332 SAVLY--CGBACLRADWQRCDDVSHRFMCPRLAHERAGERELATLPFYTALEVSET 387
 RESULT 2
 Q8NA12 PRELIMINARY; PRT; 247 AA.
 ID Q8NA12
 AC Q8NA12
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein FLJ35323.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RX PubMed=14702039;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
 RA Sudo H., Hoshiro T., Kaku Y., Kodaira H., Komori H., Sugawara M.,
 RA Takahashi N., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
 RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
 RA Nishimura K., Ishidashi T., Yamashita H., Murakami K., Fujimori K.,
 RA Tanai H., Kimata S., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
 RA Oono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
 RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
 RA Togiya S., Komai R., Hara R., Takeuchi K., Arita M., Imose N.,
 RA Musashino K., Yuki H., Oshino A., Sasaki N., Aotsuka S.,
 RA Mochikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
 RA Yoshiyama S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki Y., Watanabe K., Kumagai A., Itakura S., Fukumizu Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Taniguchi A., Fujiwara T.,
 RA Oono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
 RT CNMAs.";
 RL Nat. Genet. 36:40-45(2004).
 DR EMBL: AK092642; BAC03933.1; -
 SQ SEQUENCE 247 AA; 25975 MW; D69A7CCD02DC54C1 CRC64;

DB 223 -GSGGV-GSVGRKGM 235
 RESULT 3
 ID O17866 PRELIMINARY; PRT; 650 AA.
 AC O17866;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE CLE-1C protein (C. elegans CLE-1 protein) (Corresponding sequence
 DE C36B1.1c).
 DN Name=Cle-1; ORFNames=C36B1.1;
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21157401; PubMed=11257122;
 RA Ackley B.D., Crew J.R., Riatana H., Pihlajaniemi T., Kuo C.J.,
 RA Kramer J.M.;
 RT "The NCI/endostatin domain of Caenorhabditis elegans type XVIII
 RT collagen affects cell migration and axon guidance."
 RL J. Cell Biol. 152:1219-1232(2001).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA White S.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF164959; AA047825.1; -
 DR EMBL: Z81079; CAB03084.1; -
 DR PIR: T22002; T22002.
 DR HSSP: P39061; 1KOE.
 DR WormPep: C36B1.1C; CE17816.
 DR GO:0005737; C:cytoplasm; IEA.
 DR GO:0006817; P:phosphate transport; IEA.
 DR InterPro: IPR006160; Collagen.
 DR InterPro: IPR010515; Endostatin.
 DR Pfam: PF01391; Collagen; 2.
 DR Pfam: PF06482; Endostatin; 1.
 KW Collagen.
 SQ SEQUENCE 650 AA; 69596 MW; 6CF29ED9C16B170E CRC64;
 Query Match 11.0%; Score 93.5; DB 2; Length 650;
 Best Local Similarity 38.6%; Pred. No. 3.1;
 Matches 22; Conservative 6; Mismatches 24; Indels 5; Gaps 1;
 QY 23 PAVRLILGPPGKXKMGWGMGMSFSLPQACVXGAAPMOTGSPNDRSGQMGALGPFG 79
 DB 229 PGMGGLAGPFGPGSGSGHSGSGADPGQ-----GPGLPAGAPGDRGTSVSGCGPFG 280
 RESULT 4
 ID Q9U9K6 PRELIMINARY; PRT; 778 AA.
 AC Q9U9K6;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE CLE-1B protein (C. elegans CLE-1 protein) (Corresponding sequence
 DE C36B1.1b).
 DN Name=Cle-1; ORFNames=C36B1.1;
 OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 CX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21157401; PubMed=11257122;
 RA Ackley B.D., Crew J.R., Elamra H., Pihlajaniemi T., Kuo C.J.,
 RA Kramer J.M.;
 RT "The NCI/endostatin domain of Caenorhabditis elegans type XVIII
 RT collagen affects cell migration and axon guidance.";
 RL J. Cell Biol. 152:1219-1232(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Leonard N.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA White S.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF164959; AAD47824.1; -.
 DR EMBL; Z80215; CAD21700.1; -.
 DR EMBL; Z81079; CAD21700.1; JOINED.
 DR EMBL; Z81079; CAD21643.1; -.
 DR EMBL; Z80215; CAD21643.1; JOINED.
 DR HSSP; P39061; 1KOE.
 DR WormPep; C36B1.1b; CE29706.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008985; Conn_like_1ec_g1.
 DR InterPro; IPR010515; Endostatin.
 DR InterPro; IPR003129; TSP_N.
 DR Pfam; PF01391; Collagen; 2.
 DR Pfam; PF06482; Endostatin; 1.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00210; TSPN; 1.
 DR Collagen.
 KW SEQUENCE.
 SQ 778 AA; 83787 MW; 743AFCA7FF1ABD CRC64;
 Query Match 11.0%; Score 93.5; DB 2; Length 778;
 Best Local Similarity 38.6%; Pred. No. 3.7;
 Matches 22; Conservative 6; Mismatches 24; Indels 5; Gaps 1;
 Oy 23 PAVRLRLGPPGKAKWGWGWSFSLPQACVGAAPMOTGSPNRDGSQGMGALGPRG 79
 Db 357 FGMPLGAPGPPGSGTGHGSGADGPO-----GPPGLPGAPGRDGTSGVEGQRGPOG 408
 RESULT 5
 ID Q909K7 PRELIMINARY; PRT; 1117 AA.
 AC Q909K7;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE CLE-1A protein (C. elegans CLE-1 protein) (Corresponding sequence
 DE C36B1.1a); ORFNames=C36B1.1;
 GN Name=Cle-1; ORFNames=C36B1.1;
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 CX NCBI_TaxID=6239;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21157401; PubMed=11257122;
 RA Ackley B.D., Crew J.R., Elamra H., Pihlajaniemi T., Kuo C.J.,
 RA Kramer J.M.;
 RT "The NCI/endostatin domain of Caenorhabditis elegans type XVIII
 RT collagen affects cell migration and axon guidance.";
 RL J. Cell Biol. 152:1219-1232(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Leonard N.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA White S.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF164959; AAD47823.1; -.
 DR EMBL; Z80215; CAD21701.1; -.
 DR EMBL; Z81079; CAD21644.1; -.
 DR EMBL; Z80215; CAD21644.1; JOINED.
 DR HSSP; P39061; 1KOE.
 DR WormPep; C36B1.1a; CE29705.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008985; Conn_like_1ec_g1.
 DR InterPro; IPR010515; Endostatin.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR008957; FN_III-like.
 DR InterPro; IPR003129; TSP_N.
 DR Pfam; PF01391; Collagen; 2.
 DR Pfam; PF06482; Endostatin; 1.
 DR Pfam; PF00041; FN3; 2.
 DR SMART; SM00060; FN3; 2.
 DR SMART; SM00210; TSPN; 1.
 DR PROSITE; PS50853; FN3; 2.
 DR Collagen.
 KW SEQUENCE.
 SQ 1117 AA; 121680 MW; CF323D42E88770AD CRC64;
 Query Match 11.0%; Score 93.5; DB 2; Length 1117;
 Best Local Similarity 38.6%; Pred. No. 5.6;
 Matches 22; Conservative 6; Mismatches 24; Indels 5; Gaps 1;
 Oy 23 PAVRLRLGPPGKAKWGWGWSFSLPQACVGAAPMOTGSPNRDGSQGMGALGPRG 79
 Db 696 FGMPLGAPGPPGSGTGHGSGADGPO-----GPPGLPGAPGRDGTSGVEGQRGPOG 747
 RESULT 6
 ID Q810G6 PRELIMINARY; PRT; 1154 AA.
 AC Q810G6;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE C. elegans CLE-1 protein (Corresponding sequence C36B1.1d).
 GN Name=Cle-1; ORFNames=C36B1.1;
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 CX NCBI_TaxID=6239;

```

OX NCBI_TaxID=6239;
RN [1]
RA SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=9069613; PubMed=9851916;
RA none;
RT "genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RT Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Lemard N.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA White S.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z80215; CAD56563.1; -.
DR EMBL; Z81079; CAD56563.1; JOINED.
DR EMBL; Z81079; CAD56563.1; -.
DR EMBL; Z80215; CAD56563.1; JOINED.
DR HSSP; P39061; 1KOE.
DR WormBep; C36B1.1d; C32319.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0005198; F-structural molecule activity; IEA.
DR GO; GO:0007155; P-cell adhesion; IEA.
DR GO; GO:0006817; Phosphate transport; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008995; Coma_like_lcc_g1.
DR InterPro; IPR010515; Endostatin.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003129; TSP_N.
DR Pfam; PF06482; Collagen; 2.
DR Pfam; PF00041; fn3; 2.
DR ProDom; PD000007; C1g_helix; 1.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00853; FN3; 2.
KW Collagen.
SQ SEQUENCE 1154 AA; 125924 MW; 7B2D9B9B3CD1ABE5 CRC64;

Query Match 11.0%; Score 93.5; DB 2; Length 1154;
Best Local Similarity 38.6%; Pred. No. 5.8;
Matches 22; Conservative 6; Mismatches 24; Indels 5; Gaps 1;

QY 23 PAVLRLLGPPGKAKMGWGSFSLPQACVGAAPMGTGSPVRDSSQGMGALGPRG 79
Db 733 PGMPLAGPPPPGSGTGHSGADGPG-----GPRGLGAPGRDGTSGVEGQRPQG 784

RESULT 7
ID Q6ZNT5 PRELIMINARY; PRT; 498 AA.
AC Q6ZNT5;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein FLJ16083.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Houta T., Hiseao S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,

```

```

RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Nagatsuma M., Takahashi-Fujii A., Oshina A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.,
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131209; BADI8400.1; -.
DR InterPro; IPR007087; ZnF_C2H2.
DR Pfam; PF00096; zf-C2H2; 14.
DR ProDom; PD000003; ZnF_C2H2; 12.
DR SMART; SM00355; ZnF_C2H2; 14.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 14.
SQ SEQUENCE 498 AA; 54712 MW; 7C662E7C22946B0F CRC64;

Query Match 10.7%; Score 90.5; DB 2; Length 498;
Best Local Similarity 31.1%; Pred. No. 4.4;
Matches 42; Conservative 7; Mismatches 39; Indels 47; Gaps 11;

QY 33 GKAKMGW-WGWSFSLP-----QACVGAAPMGTGSPNRD--GSGMDGALGPRGW 80
Db 32 GAVSGGAWGAMENSTVEPRAGDQGRQATLGAADGQ-GGPRGLPADGGRDGA-GPRS- 88

QY 81 TPASCHLPLRQSALHPSPSLIYESIGVC-----VTTT-----PYRCPSN 120
Db 89 EPA-----DRALRPP--LPBPGRCCGCGKASQGSYLQHRVRHTGKPYTCPEC 139

QY 121 TQHYHLRPNSELSQ 135
Db 140 GKAPAW---SSNLSQ 151

RESULT 8
ID Q6ZTD2 PRELIMINARY; PRT; 498 AA.
AC Q6ZTD2;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein FLJ44773.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Cerebellum;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Mitsuhashi K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Saito H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Nagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.,
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK126727; BAC86660.1; -.
DR InterPro; IPR007087; ZnF_C2H2.
DR Pfam; PF00096; zf-C2H2; 14.
DR ProDom; PD000003; ZnF_C2H2; 11.
DR SMART; SM00355; ZnF_C2H2; 14.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 14.
SQ SEQUENCE 498 AA; 54652 MW; 7C762F70538B1B0F CRC64;

Query Match 10.7%; Score 90.5; DB 2; Length 498;
Best Local Similarity 31.1%; Pred. No. 4.4;
Matches 42; Conservative 7; Mismatches 39; Indels 47; Gaps 11;

QY 33 GKAKMGW-WGWSFSLP-----QACVGAAPMGTGSPNRD--GSGMDGALGPRGW 80
Db 32 GAVSGGAWGAMENSTVEPRAGDQGRQATLGAADGQ-GGPRGLPADGGRDGA-GPRS- 88

QY 81 TPASCHLPLRQSALHPSPSLIYESIGVC-----VTTT-----PYRCPSN 120

```

Db 89 EPA-----DRALRSP--LPBPGRCGCGKAFSGSYLLQHRVHTGKPYTCPEC 139

QY 121 TOHYHWLRPNSELSQ 135
: : : : :
Db 140 GKAFAM---SSNLSQ 151

RESULT 9
BAC6660 PRELIMINARY; PRT; 498 AA.

AC BAC6660:
DT 02-MAR-2004 (TREMblrel. 27, Created)
DT 02-MAR-2004 (TREMblrel. 27, Last sequence update)
DE CDNA FLJ44773 fis, clone BRACE3034183, moderately similar to zinc finger protein 84.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K., Arita M., Mueashiro K., Yuuki H., Hara H., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Nagatsuna M., Sugawara K., Kaneshori K., Sugiyama A., Kawakami B., Suzuki Y., Sano S., Nagahari K., Masuo Y., Nagai K., Isogai T., "NEDO human cDNA sequencing project."
RT Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK16727; BAC6660.1; -
SQ SEQUENCE 498 AA; 54652 MW; 7C762F7053881B0F CRC64;

Query Match 10.7%; Score 90.5; DB 2; Length 498;
Best Local Similarity 31.1%; Pred. No. 4.4;
Matches 42; Conservative 7; Mismatches 39; Indels 47; Gaps 11;

QY 33 GKAKWGC-WGWSFSLP-----QACVXGAAPMOTGSPNDR--GSQMDGALGPRGM 80
Db 32 GAVSGGCGAMENSTEVPRAGDQGRQATLGAADQ--GGPGRLEPAGDGRGA-GPRS- 88
QY 81 TPASCHLPLRQALHPSPSLIYESIGSYC-----VTH-----PYRCPN 120
Db 89 EPA-----DRALRSP--LPBPGRCGCGKAFSGSYLLQHRVHTGKPYTCPEC 139

QY 121 TOHYHWLRPNSELSQ 135
: : : : :
Db 140 GKAFAM---SSNLSQ 151

RESULT 10
BADI8400 PRELIMINARY; PRT; 498 AA.

AC BADI8400:
DT 12-MAY-2004 (TREMblrel. 27, Created)
DT 12-MAY-2004 (TREMblrel. 27, Last sequence update)
DE CDNA FLJ16083 fis, clone NT2R12008952, moderately similar to ZINC FINGER PROTEIN 84.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum;
RA Ishida T., Kaneshori K., Yosida M., Watanabe S., Ishida S., Ono Y., Houta T., Hitaoka S., Murakawa K., Takiguchi S., Kusano J., Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,

RA Wagatsuna M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuo Y., Nagai K., Isogai T., "NEDO human cDNA sequencing project."
RT Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK131209; BADI8400.1; -
SQ SEQUENCE 498 AA; 54712 MW; 7C662E7C22946B0F CRC64;

Query Match 10.7%; Score 90.5; DB 2; Length 498;
Best Local Similarity 31.1%; Pred. No. 4.4;
Matches 42; Conservative 7; Mismatches 39; Indels 47; Gaps 11;

QY 33 GKAKWGC-WGWSFSLP-----QACVXGAAPMOTGSPNDR--GSQMDGALGPRGM 80
Db 32 GAVSGGCGAMENSTEVPRAGDQGRQATLGAADQ--GGPGRLEPAGDGRGA-GPRS- 88
QY 81 TPASCHLPLRQALHPSPSLIYESIGSYC-----VTH-----PYRCPN 120
Db 89 EPA-----DRALRSP--LPBPGRCGCGKAFSGSYLLQHRVHTGKPYTCPEC 139

QY 121 TOHYHWLRPNSELSQ 135
: : : : :
Db 140 GKAFAM---SSNLSQ 151

RESULT 11
Q9J104 PRELIMINARY; PRT; 1737 AA.

AC Q9J104:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE Alpha 4 type V collagen.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague Dawley;
RX MEDLINE=20428740; PubMed=10852920;
RA Chernousov M.A., Rothblum K., Tyler W.A., Stahl R.C., Carey D.J., "Schwann cells synthesize type V collagen that contains a novel alpha RT affinity heparin binding of alpha 4(V) collagen."
RT J. Biol. Chem. 275:28208-28215 (2000).
RL EMBL; AF272651; AAF76432.1; -
HSSP; P04278; 1D2S.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cyclopasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; collagen.
DR InterPro; IPR008985; CnA_like_lcc_g1.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR003129; TSP_N.
DR Pfam; PF01410; COLF1; 1.
DR Pfam; PF01391; Collagen; 17.
DR Pfam; PF02210; laminin_G_2; 1.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
KW Collagen.
SQ SEQUENCE 1737 AA; 171574 MW; D635D57481C257 CRC64;

Query Match 10.4%; Score 88; DB 2; Length 1737;
Best Local Similarity 25.3%; Pred. No. 30;
Matches 46; Conservative 19; Mismatches 51; Indels 66; Gaps 11;

QY 25 VRLIGPPGKAKKWKMGWSFSLPQACVXGAAPMOTGSPNDR--GSQMDGALGPRGM 64
: : : : :
Db 1404 LIGLIGPPGGA---GEKDGQJFG--VGGPGLQDPLGPGVSGLGHGPPGVVPLG 1457

QY 65 RDGSGMDGALGPRG-----WTPASCH-LPFRGSLH-PSBSL--IYESIGSYC 109
 Db 1458 QKSGKSGPSTLGRGDPGAPGPPGPPSPAEVHGLRRRRSVLTPGCGLEEVASINIS 1517
 QY 110 VTHPRYCPENT-----QHYH-----WLRNPSLSQIRFTISIFIN---G 146
 Db 1518 LELFQLRPPGTABSPGLICHELHRNPHLPDGEYWDPNQCARAF--KVFCHPTAGG 1575
 QY 147 ET 148
 Db 1576 ET 1577

RESULT 12
 ID CA35_HUMAN STANDARD; PRT; 1745 AA.
 AC P25940; Q9NZ06;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Collagen alpha 3 (V) chain precursor.
 GN Name=COL5A3;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart, and Placenta;
 RX MEDLINE=20187594; PubMed=10722718;
 RA Imanura Y., Scott I.C., Greenspan D.S.;
 RT "The pro-alpha 3 (V) collagen chain, Complete primary structure,
 RT expression domains in adult and developing tissues, and comparison to
 RT the structures and expression domains of the other types V and XI
 RT procollagen chains.";
 RL J. Biol. Chem. 275:8749-8759 (2000).
 RN [2]
 RP PRELIMINARY SEQUENCE OF 479-564; 665-709; 723-758; 787-816; 922-1008;
 RP 1054-1088; 1248-1287 AND 1313-1334.
 RC TISSUE=Placenta;
 RX MEDLINE=92239022; PubMed=1571108;
 RA Mann K.;
 RT "Isolation of the alpha 3-chain of human type V collagen and
 RT characterization by partial sequencing.";
 RL Biol. Chem. Hoppe-Seyler 373:69-75 (1992).
 CC -1- FUNCTION: Type V collagen is a member of group I collagen
 CC (fibrillar forming collagen). It is a minor connective tissue
 CC component of nearly ubiquitous distribution. Type V collagen binds
 CC to DNA, heparan sulfate, thrombospondin, heparin, and insulin.
 CC -1- SUBUNIT: Trimers of two alpha 1 (V) and one alpha 2 (V) chains in
 CC most tissues and trimers of one alpha 1 (V), one alpha 2 (V), and
 CC one alpha 3 (V) chains in placenta.
 CC -1- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- SIMILARITY: Belongs to the fibrillar collagen family.
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL substation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AF177941; AA059902.1; -
 DR EMBL; S20375; S20375
 DR Genew; HGNC;14864; COL5A3.
 DR MIM; 120216; -
 DR GO; GO:0005588; C:collagen type V; TAS.
 DR GO; GO:0005201; Extracellular matrix structural constituent; NAS.
 DR InterPro; IPR008161; C1g helix.
 DR InterPro; IPR008160; Collagen.

DR InterPro; IPR008985; Cona like lec gl.
 DR InterPro; IPR008985; Fib collagen_C.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR003129; TSP_N.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 17.
 DR Pfam; PF02210; TSP_N; 1.
 DR Prodom; PD000007; C1g helix; 2.
 DR Prodom; PD002078; Fib collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00282; LamC; 1.
 DR SMART; SM00210; TSPN; 1.
 KW Collagen; Connective tissue; Direct protein sequencing;
 KW Extracellular matrix; Hydroxylation; Repeat; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 1745 Collagen alpha 3 (V) chain.
 FT DOMAIN 31 210 TSP N-terminal.
 FT DOMAIN 211 381 Nonhelical region.
 FT DOMAIN 392 1489 Triple-helical region.
 FT DOMAIN 1490 1745 Nonhelical region.
 SQ SEQUENCE 1745 AA; 172051 MW; 5B8FF97135397AC1 CRC64;

Query Match 10.4%; Score 88; DB 1; Length 1745;
 Best local similarity 24.2%; Pred. No. 30;
 Matches 45; Conservative 17; Mismatches 54; Indels 70; Gaps 9;

QY 25 VLRLGPPGAKKMGWNGMSFSLPQACVXG-----AAPMGTGSPN 64
 Db 1407 LIGLIGPPGEX-----GEKDGQLFG--VQPPGPKGDPGPPGIGLGHPPGVPGLG 1460
 QY 65 RDGSGMDGALGPRG-----WTPASCH-LPFRGSLH-PSBSL--IYESIGSYC 109
 Db 1461 QKSGKSGPSTLGRGDPGAPGPPGPPSPAEVHGLRRRRSVLTPGCGLEEVASINIS 1517
 QY 106 GSVCTVTHPRYCPENT-----QHYH-----WLRNPSLSQIRFTISIFIN- 145
 Db 1521 TSLSLLEQLRRPPTABSPGLICHELHRNPHLPDGEYWDPNQCARAF--KVFCHPTAGG 1578
 QY 146 ---GET 148
 Db 1579 TAGGET 1584

RESULT 13
 ID Q9H2C6 PRELIMINARY; PRT; 1078 AA.
 AC Q9H2C6;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Cellulose synthase.
 GN Name=CesA5;
 OS Populus tremuloides (Quaking aspen).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Malpighiales; Salicaceae; Populus.
 OX NCBI_TaxID=3693;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Joshi C.P., Kalluri U.C.;
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY055724; AA23710.2; -
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0016760; F:cellulose synthase (UDP-forming) activity; IEA.
 DR GO; GO:0030244; P:cellulose biosynthesis; IEA.
 DR InterPro; IPR005150; Cellulose synth.
 DR Pfam; PF03552; Cellulose synth; 1.
 SQ SEQUENCE 1078 AA; 119539 MW; 3CDA85C184D7B53F CRC64;


```

Query Match      10.3%; Score 87.5; DB 2; Length 1076;
Best Local Similarity 22.2%; Pred. No. 20;
Matches 39; Conservative 23; Mismatches 65; Indels 49; Gaps 8;

QY 19 LHETPAV-----LRLGPPGKAKWGMG---GMSFSLPQACVXGAAPMOTG-SPNRDGSQ 69
DB 178 VHGSNNRVVDPVPAEFSGPGLGVAMKERVYDGMKQDKTVV-----PMSTGHAPSERGAG 233
QY 70 GMDGA-----LGRGWTTPASCHPLRQSAHHPSPSLYBSISGVCTTHPRCPSPN 120
DB 234 DIDATVDVLDSDSLNDEARQPLSRKVSIPESKINPRMIVLGLGICFLH-YRITNP 292
QY 121 TQHH-----WLRPNSLSQIRFTISIFINGSETSPSH 152
DB 293 VRNAVAGLVIGMGDMFAISRIIDQPPKMLPKNRETYLDRLALRYMBEGE--PSH 346

RESULT 14
Q6NZ15 PRELIMINARY; PRT; 1449 AA.
ID Q6NZ15
AC Q6NZ15
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Collagen, type I, alpha 3.
GN Name=coll1a3;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22386257; PubMed=12477932;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Strausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood U., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzywinski M.I., Skalska U., Smaltz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Strausberg R.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC066384; AAH66384.1; -
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR000885; Fibrinogen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 17.
DR ProDom; PD000007; Clg_helix; 4.
DR ProDom; PD002078; Fibrinogen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.

```

```

DR PROSITE; PS01208; VWC_1; 1.
DR PROSITE; PS0184; VWC_2; 1.
KW Collagen.
SQ SEQUENCE 1449 AA; 136980 MW; A763D6C12BDFD07 CRC64;

Query Match      10.2%; Score 86.5; DB 2; Length 1449;
Best Local Similarity 22.6%; Pred. No. 34;
Matches 44; Conservative 12; Mismatches 54; Indels 85; Gaps 9;

QY 30 GPPG-KAKGKGMGMSFSLPQACVXGAAPM-----QTGSPNRGSGGMSGALGPRG----- 79
DB 1108 GPPGTSGEQG-----PAGATPAGRGPGAGSNGSGKGNKGVGPVGPGRGAS 1158
QY 80 -----WTPASCHPLRQSAHHPSP-----SLIYE 103
DB 1159 GEMGPAGAPGLPGPPPGPSAGPGEPLFMPQEKGPDPRLRGYRADASVVRDSEVD 1218
QY 104 SIGVCVTHPRPCPENTQ-----HYWLRPN--SELSQIRFTISI 142
DB 1219 SLKSLSQKTIENIRSPGTOANPARMCRDLRMCHPPEWKSQSVWVDNQGSLDAIK-----V 1274
QY 143 FINGER-----SPSH 152
DB 1275 FCNMTGFTCVSDPYH 1289

RESULT 15
Q6PE19 PRELIMINARY; PRT; 1449 AA.
ID Q6PE19
AC Q6PE19
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Collagen, type I, alpha 3.
GN Name=coll1a3;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22386257; PubMed=12477932;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Strausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood U., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzywinski M.I., Skalska U., Smaltz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Strausberg R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC058045; AAH58045.1; -
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR002181; Fibrinogen_C.

```

DR InterPro: IPR000885; Fib_collagen_C.
DR InterPro: IPR001007; WMF_C.
DR Pfam: PF04410; COLFI, 1.
DR Pfam: PF01391; Collagen; 17.
DR Pfam: PF00093; WWC; 1.
DR ProDom: PD000007; C1g_helix; 4.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR SMART: SM00214; WWC; 1.
DR PROSITE: PS01208; WWC_1; 1.
DR PROSITE: PS50184; WMF_2; 1.
DR Collagen.
SQ SEQUENCE 1449 AA; 136930 MW; 9A51BA0B674B08B8 CRC64;

Query Match 10.2%; Score 86.5; DB 2; Length 1449;
Best Local Similarity 22.6%; Pred. No. 34;
Matches 44; Conservative 12; Mismatches 54; Indels 85; Gaps 9;

QY 30 GPRG-KAKMGWGWMSFSLPQACVXGAAPN---QTGSPNRDGSQGMGALGPRG----- 79
Db 1108 GPRGTSGEQG-----PAGATGPAGPRGPAAGNGSPGKDGAMGVPGVGPGRGRS 1158
QY 80 -----WTPASCHLPLRQSLHPS-----SLAYE 103
Db 1159 GEMGPAGAPGLPGPPGPGSAGPGEPPLPMPQOEKGDPLRGYRADNASVRDRDSEVDT 1218
QY 104 SIGSVCTTHPYRCPSNTQ-----HYHMLRPN--SELSQIRFTISI 142
Db 1219 SLKSLSGKIENTIRSPGCTQANPARMCRLDMCHPEWKSQSYWVDPNQGSPLDAIK-----V 1274
QY 143 FINGET-----SPSH 152
Db 1275 FCNNETGETCVSPYH 1289

Search completed: October 24, 2004, 16:03:24
Job time : 97 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OK protein - nucleic search, using frame_plus_p2n model

Run on: October 24, 2004, 16:03:58 ; Search time 2704 Seconds
(without alignments)
2048.387 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848
Sequence: 1. MPHLYIDGVFPPIQVREKLN.....LSQIRFTISIFINGETSPSH 152

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h
-Q=/cgn2_1/USPTO.spool/US09977418/runat_22102004_115948_1031/app_query.fasta_1.327
-DB=EST -QFMT=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blowm62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USERS=US09977418 @CCN 1.1_3437 @runat_22102004_115948_1031 -NCPD=6 -ICPU=3
-NO_WMAP -LARGEDUTRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: gb_est1:
2: gb_est2:
3: gb_est3:
4: gb_est4:
5: gb_est5:
6: gb_est6:
7: gb_est7:
8: gb_est8:
9: gb_est9:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	817	96.3	901	7	CF552903 AGNCOURT
2	777	91.6	1139	4	BM926609 AGNCOURT
3	765	90.2	974	2	BM250170 60094555
4	753	88.8	508	6	CB267862
5	752	88.7	878	5	EX402471 BX402471
6	748	88.2	726	2	BF792513 60225622
7	731	86.2	624	2	W61026
8	709	83.6	636	2	BF306561 60188986
9	683	80.5	496	1	AJ707312

10	666	78.5	376	6	CB267751	1006657 H
11	662	78.1	528	1	AJ711705	AJ711705
12	662	78.1	956	2	BF203383	60186595
13	662	78.1	976	2	BF203719	60186380
14	631	74.4	518	6	CD625679	56050576
15	583	68.8	562	4	BI113724	60286019
16	531	62.6	848	7	CF552003	CF552003
17	527	62.1	605	2	AM384208	RCO-HT037
18	522	61.6	416	1	AJ711115	AJ711115
19	506	59.7	951	2	BF306631	60188886
20	494	58.3	501	1	AJ711059	AJ711059
21	487	57.4	443	2	BE840664	RC1-FN018
22	479.5	56.5	508	7	N55874	U3561F Huma
23	468	55.2	503	1	AA211656	AA211656 zn56e03.1
24	462	54.5	409	2	BE840591	RC1-FN018
25	456	53.8	400	6	CD625675	56000825
26	456	53.8	400	6	CD625677	56000901
27	449	52.9	404	6	CD625674	56000825
28	449	52.9	409	6	CD625676	56000901
29	436	51.4	502	1	AJ708694	AJ708694
30	428.5	50.5	400	6	CD625678	56050576
31	418	49.3	389	6	CD4005	CD4005
32	396	46.7	227	1	AA042804	AA042804
33	372	43.9	395	6	CB050516	CB050516
34	372	43.9	395	6	CB050517	NISC-9118
35	372	43.9	395	6	CB050517	NISC-9118
36	330	38.9	690	5	BC006459	BC006459
37	330	38.9	690	5	BM936902	UI-H-ED0-
38	322	38.0	329	1	AI805232	AI805232
39	321	37.9	329	1	AI817452	AI817452
40	321	37.9	329	1	AI817826	AI817826
41	321	37.9	329	1	AI817826	AI817826
42	318	37.5	533	1	AJ709557	AJ709557
43	306	36.1	183	1	AJ709804	AJ709804
44	301	35.5	236	7	F00226	HS886H051 S
45	298	35.1	672	6	CA424992	UI-H-FE1-

ALIGNMENTS

RESULT 1
CF552903
LOCUS
AGNCOURT 15596050 NIH MGC 183 Homo sapiens CDNA clone
IMAGE:10529648 5', MENA sequence.

CF552903
VERSION
CF552903.1 GI:34889737
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
NIH-MGC http://mgs.nci.nih.gov/.
1 (baee 1 to 901)
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgaabbs-remail.nih.gov

Tissue Procurement: Dr. Michael Brownstein
CDNA Library Preparation: Invitrogen Corp
DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNI at:
http://image.llnl.gov
Plate: NDAM617 row: 1 column: 17
High quality sequence stop: 674.
Location/Qualifiers
1..901
/organism="Homo sapiens"

```
/mol_type="mrna"  
/db_xref="taxon:9606"  
/clone="IMAGE:30529648"  
/lab_host="DH10B-nona (T1 and T5 phage resistant)"  
/clone_lib="NIH_MGC_183"  
/note="Organ: Pooled muscle (cardiac and skeletal);  
Vector: PCMV-Sports.1; Site_1: EcoRV (destroyed); Site_2:  
NotI; Library is oligo-dT primed and directionally cloned  
(EcoRV site is destroyed upon cloning). Average insert  
size 1.7. Library was constructed by Invitrogen."
```

ORIGIN

Alignment Scores:

Pred. No.:	1,766-57	Length:	901
Score:	817.00	Matches:	150
Percent Similarity:	98.05%	Conservative:	1
Best Local Similarity:	97.40%	Mismatches:	1
Query Match:	96.34%	Indels:	2
DB:	7	Gaps:	0

US-09-977-418-8 (1-152) x CF552903 (1-901)

```
QY 1 MetProHisLeuTyrlleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20  
DB 231 ATGCCACATCTGTATATAGATGGGCTTTTCCAAATACAGCTGTCGTGATTAACCTGCAT 290  
QY 21 GluThrProAlaValleuArgLeuGlyProProGlyLysAlaLysTrpGlyTrpGly 40  
DB 291 GAAACTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 350  
QY 41 TrpGlyTrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60  
DB 351 TGGGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 410  
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTrp 80  
DB 411 GGATCACCCTAACAGAGATGGAGCCAGGCGCATGATGAGGCTTGGGCTCTCGAGGTTGG 470  
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerle 100  
DB 471 ACCCCAGCTTCTTGCCACCTTCCCTCCCGGCGAGTGTAGCTTCCATCCATCCCTCTTTT 530  
QY 100 uileTyrgLuserIleGlySerValCysValThr-ThrHisProTyrgCysProserA 120  
DB 531 AATCTATGATCATATAGGCTCGGTGTGTGTAACACACACACACCTATGCTTCTCTTCA 590  
QY 120 AsnThrGlnHisTyrlleHisTrpLeuArgProAsnSerGlnLeuSerGlnLeuArgPheTrp 140  
DB 591 AATCTAGACATTAACCATGGTGTGAGGCGCAATTCAAGAGCTTCTCAATCAGATTACAA 650  
QY 140 leSerIlePheIleAsnGlyGluThrSerProSerHis 152  
DB 651 TCTCCATTTTTCATTACGGGGAACATCCCGAGCCAC 688
```

```
RESULT 2  
BM926609 1139 bp mRNA linear EST 12-MAR-2002  
LOCUS AGENCOURT 6644683 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5767017  
DEFINITION 5', mRNA sequence.  
ACCESSION BM926609  
VERSION BM926609.1 GI:19376988  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1139)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.
```

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LNL412824 row: j column: 10
High quality sequence stop: 717.
Location/Qualifiers
1. 1139
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="IMAGE:5767017"
/lab_host="DH10B"
/clone_lib="NIH_MGC_122"
/note="Organ: pooled lung and spleen; Vector: PCMV-Sports6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source:
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.:	4,866-54	Length:	1139
Score:	777.00	Matches:	148
Percent Similarity:	95.51%	Conservative:	1
Best Local Similarity:	94.87%	Mismatches:	3
Query Match:	91.63%	Indels:	4
DB:	4	Gaps:	0

US-09-977-418-8 (1-152) x BM926609 (1-1139)

```
QY 1 MetProHisLeuTyrlleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20  
DB 398 ATGCCACATCTGTATATAGATGGGCTTTTCCAAATACAGCTGTCGTGATTAACCTGCAT 457  
QY 21 GluThrProAlaValleuArgLeuGlyProProGlyLysAlaLysTrpGlyTrpGly 40  
DB 458 GAAACTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 517  
QY 41 TrpGlyTrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60  
DB 518 TGGGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 577  
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTrp 80  
DB 578 GGATCACCCTAACAGAGATGGAGCCAGGCGCATGATGAGGCTTGGGCTCTCGAGGTTGG 637  
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerle 100  
DB 638 ACCCCAGCTTCTTGCCACCTTCCCTCCCGGCGAGTGTAGCTTCCATCCATCCCTCTTTT 697  
QY 100 uileTyrgLuserIleGlySerVal-CysValThr-ThrHisProTyrgCysProser 119  
DB 698 AATCTAGACATTAACCATGGTGTGAGGCGCAATTCAAGAGCTTCTCAATCAGATTAC 757  
QY 120 AsnThrGlnHisTyrlleHisTrpLeuArgProAsnSerGlnLeuSerGlnLeuArgPheTrp 139  
DB 758 AATCTAGACATTAACCATGGTGTGAGGCGCAATTCAAGAGCTTCTCAATCAGATTAC 817  
QY 139 IleSerIlePheIleAsnGlyGluThrSerProSerHis 152  
DB 818 AATCTCCATTTTTCATTACGGGGAACATCCCGAGCCAC 857  
RESULT 3  
BE250170 974 bp mRNA linear EST 13-JUL-2000  
LOCUS BE250170
```

DEFINITION 60094355F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2960076 5', mRNA sequence.

ACCESSION BE250170

VERSION BE250170.1 GI:9120277

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 974)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: gsapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LUCM52 Row: n Column: 13
High quality sequence stop: 692.
Location/Qualifiers

FEATURES

source 1..974

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2960076"

/tissue_type="rhabdomyosarcoma"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC 17"

/note="Organ: muscle; Vector: pOTB7; Site 1: EcoRI; Site 2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected by for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:

Pred. No.: 3,89e-53 Length: 974

Score: 765.00 Matches: 147

Percent Similarity: 95.48% Conservative: 1

Best Local Similarity: 94.84% Mismatches: 4

Query Match: 90.21% Indels: 4

DB: 2 Gaps: 0

US-09-977-418-8 (1-152) x BE250170 (1-974)

QY 1 MePProHsiLeuYrTtLeaSGlYValPheProIleGlnLeuValArgGlnUlysLeuHs 20

DB 211 ATGCCAATCTGATATAGATGAGGTTTTCATTAACAAGCTGCTCTGATATACATGCA 270

QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTTPGly 40

DB 271 GAAACTCTGCGCGCTCGCGCTGCTGAGGCTCCAGCAAGGCAAGTGGGGTGGGG 330

QY 41 ThrGlyTrpSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThr 60

DB 331 TGGGGCTGGTCTCTCTCCCTCCACAGGCTGtTCTTGGGGCTGCTCCACAGGACACA 390

QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTTP 80

DB 391 GGATCACTACACAGATGAGAACCA-GGCATGATGGGGCTTGGGCTCTCGAGGTGG 449

QY 81 ThrTrpAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100

DB 450 ACCCAAGCTTCTTCCACACTTCCCTCCGAGCATCGCTCTCCATCCATCCCTCTTT 509

QY 100 uileTyrGlnUserIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA 120

DB 510 AATCTATGATCTAATAGGCTCGTGTGTGTAAACACACACCCCTATGCTGTCTTCA 569

QY 120 snThrGlnHisTyrHisTrpLeuArgProAsnSer-GluLeuSerGlnIleArgPheThr 139

DB 570 AATCTCAGATTAACCATTTGGTTGAGGCCAATCAAGAGCTTCTCAATCAGATTACA 629

QY 140 IleSerIlePheIleAsnGlyGluThrSerProSerHis 152

DB 630 ATCTCATTTTCATTACAGGGGAAAAACATCCGAGCCAC 668

RESULT 4

LOCUS CB267862

DEFINITION 1006768 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens

ACCESSION CB267862

VERSION CB267862.1 GI:28442447

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Yang,R.-Z., Shuldiner,A. and Gong,D.-W. 1 (bases 1 to 508)

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE EST analysis of human adipose gene expression

JOURNAL Unpublished (2002)

COMMENT Contact: Gong Da-Wei
Division of Endocrinology, Diabetes and Nutrition
University of Maryland
660 Redwood St, HH497, Baltimore, MD 21201, USA
Tel: 410 706 1672
Fax: 410 706 1622
Email: dgong@medicine.umaryland.edu
PCR Primers
FORWARD: CTCGGGAAGCGGCCATTGTGTGTT
BACKWARD: AATACGACTACATATAGGCGCAATTTGG
Seq primer: GTTGGTACCCGGGAATTC.
Location/Qualifiers

FEATURES

source 1..508

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/sex="Male and Female"

/tissue_type="Adipose"

/clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"

/note="Vector: lambdaTriplEx"

ORIGIN

Alignment Scores:

Pred. No.: 1.63e-52 Length: 508

Score: 753.00 Matches: 139

Percent Similarity: 97.90% Conservative: 1

Best Local Similarity: 97.20% Mismatches: 1

Query Match: 88.80% Indels: 2

DB: 6 Gaps: 0

US-09-977-418-8 (1-152) x CB267862 (1-508)

QY 12 IleGlnLeuValArgGlnUlysLeuHisGlnThrProAlaValLeuArgLeuGlyPro 31

DB 1 AATCAGCTGCTTCGATTAAGTGAATGATTAAGTCTCTCCGCTCTGCTGAGGCTCT 60

QY 32 ProGlyLysAlaLysTTPGlyTTPGlyTTPGlyTrpSerPheSerLeuProGlnAlaCys 51

DB 61 CCAAGCAAGGCCAATGGGGTGGGGCTGGGGCTGCTCTCTCTCCCTCCACAGGCTGT 120

QY 52 Val***GlyAlaAlaProMetGlnThrGlySerProAsnArgAspGlySerGlnGlyT 71

DB 121 GTTCTTGGGGCTGCTCCATGCAAGACAGATCACTTAACAGATGAGAACCCAGGCA 180

QY 72 AspGlyAlaLeuGlyProArgGlyTTPThrProAlaSerCysHisLeuProLeuArg-G 91

Db	18	GATGGGGCTTTGGTGCTCTGAGGTGAGACCCAGAGTCTTGCCACCTCCCTCCGGGCA	240
Qy	91	nsrAlaLeuHisProSerProSerLeuIleTyrGluSerIleGlySerAlaCysValTh	111
Db	241	GTAGGTCTCCATCCATCCCTCTTTATATATGAATCATAGGCTGGTGATGTAAC	300
Qy	111	r-ThrHisProTyrArgCysProSerAsnTrpGlnHisTyrHisTyrIleArgProAsn	131
Db	301	ACACACACCCCTATCTGTTCTCTTCAATTAAGCATGATTCACATTCAGTGGTGAAGCAAA	360
Qy	131	ergLeuLeuSerGlnIleArgPheTrpIleSerIlePheIleAsnGlyGluThrSerPro	151
Db	361	CAGAGCTTCTCATAAATCAGATTTCACAACTTCATTTCATTAAACGGGAGAAATCCCGCA	420
Qy	151	erHis	152
Db	421	GCCAC	425
RESULT 5			
LOCUS	EX402471	878 bp	mRNA
DEFINITION	EX402471 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA		linear
ACCESSION	clone CS00D039YB13	3-PRIME	mRNA sequence.
VERSION	EX402471		
KEYWORDS	EX402471.2	GI:46874966	
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLES	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
JOURNAL	1 (bases 1 to 878)		
COMMENT	Full-length cDNA libraries and normalization		
	Unpublished (2001)		
	On May 13, 2003 this sequence version replaced gi:30618825.		
	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	BP 181 91006 Evry cedex - France		
	Email: seq@genoscope.cns.fr, Web : www.genoscope.cns.fr		
	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime		
	end enriched, double-strand cDNA was digested with Not I and cloned		
	into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library		
	was normalized. Library was constructed by Life Technologies, a		
	division of Invitrogen. This sequence belongs to sequence cluster		
	545.1		
	For more information about this cluster, see		
	http://www.genoscope.cns.fr/cdnats-CS1A1010ZF07NP1&c=545.1.		
FEATURES	Location/Qualifiers		
SOURCE	1..878		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CS00D039YB13"		
	/issue_type="PLACENTA COT 25-NORMALIZED"		
	/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"		
	/note="1st strand cDNA was primed with a NotI-oligo(dT)		
	primer. Five prime end enriched, double-strand cDNA was		
	digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."		
ORIGIN			
Alignment Scores:			
Pred. No.:	4,04e-52	Length:	878
Score:	752.00	Matches:	135
Percent Similarity:	94.44%	Conservative:	1
Best Local Similarity:	93.75%	Mismatches:	8
Query Match:	88.68%	Indels:	0
DB:	5	Gaps:	0
US-09-977-418-8 (1-152) x EX402471 (1-878)			
Qy	9	valpheproillegluleuvalargclulysleuistglunhprproalvalleuargleu	28

Db	876	GTTTTCCAAACACAGCTGTTCCGATTAACATGATGAACACTCCTGCCGCTCTGCGCTG	817		
Qy	29	LeuGlyProPocGlyValAlaValTrrGlyTrrGlyTrrGlyTrrPseSerPseSerLeuPro	48		
Db	816	CTGGGGCCCTCAGGCCAAGCCCAATGGAGTGGGGAGTGGGGAGTGGGGCTGTCTCTCCCTCCCA	757		
Qy	49	GlnAlaCysVal**GlyAlaAlaProMetGlnThrGlySerProAsnArgAspGlySer	68		
Db	756	CAGGCTGTGTCTTGGGGGCTGGTCCCATGAGACAGAGATCACTTAACAGATGAGAAGC	697		
Qy	69	GlnGlyMetLeuPocGlyAlaLeuGlyProArgGlyTrrPheProAlaSerCysHisLeuPro	88		
Db	696	CAGGGCATGTGATGGGGGCTTTGGGCTTCGAGGTTGAGCCCAAGCTTCTGGCACCTTCCC	637		
Qy	89	LeuArgGlnSerAlaLeuHisIspProSerProSerLeuLeuTrrGlnSerIleGlySerVal	108		
Db	636	TCCGGGCAAGTACAGCTCTCCATCCATCCCTCCCTTTATCATATGATCATATAGCTCGGT	577		
Qy	109	CysValThrThrHisIspTrrArgCysProSerAsnThrGlnHisTrrHisTrrPleuArg	128		
Db	576	GTGTACACACACACCCCTATCGTGTTCCTCAATATCTCAGCATTCATCCATGGTTAGG	517		
Qy	129	ProAsnSerGlnLeuSerGlnIleArgPheThrIleSerIlePheIleAsnGlyGlyThr	148		
Db	516	CCAAATTCAGAGCTTTCCTCAATATGATTTTACATTCCTCATTTTCATTAAAGGGGAACA	457		
Qy	149	SerProSerHis	152		
Db	456	TCCCCGAGCCAC	445		
REFERENCE	BF792513	726 bp	mrna	linear	EST 12-JAN-2001
LOCUS	602253622F1	NIH_MGC_84	Homo sapiens	CDNA clone	IMAGE:4345876 5,
DEFINITION	BF792513	BF792513.1	GI:120975567	EST	
ACCESSION	BF792513	BF792513	GI:120975567	EST	
VERSION	BF792513	BF792513	GI:120975567	EST	
KEYWORDS	BF792513	BF792513	GI:120975567	EST	
SOURCE	BF792513	BF792513	GI:120975567	EST	
ORGANISM	BF792513	BF792513	GI:120975567	EST	
REFERENCE	BF792513	BF792513	GI:120975567	EST	
AUTHORS	BF792513	BF792513	GI:120975567	EST	
TITLE	BF792513	BF792513	GI:120975567	EST	
JOURNAL	BF792513	BF792513	GI:120975567	EST	
COMMENT	BF792513	BF792513	GI:120975567	EST	
FEATURES	BF792513	BF792513	GI:120975567	EST	
Source	BF792513	BF792513	GI:120975567	EST	
1..726	BF792513	BF792513	GI:120975567	EST	
/organism="Homo sapiens"	BF792513	BF792513	GI:120975567	EST	
/mol_type="mrna"	BF792513	BF792513	GI:120975567	EST	
/db_xref="taxon:9606"	BF792513	BF792513	GI:120975567	EST	
/clone="IMAGE:4345876"	BF792513	BF792513	GI:120975567	EST	
/tissue_type="adrenal cortex carcinoma, cell line"	BF792513	BF792513	GI:120975567	EST	
/lab_host="DH10B (phage-resistant)"	BF792513	BF792513	GI:120975567	EST	
/clone_1lb="NIH_MGC_84"	BF792513	BF792513	GI:120975567	EST	
/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dt primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies."	BF792513	BF792513	GI:120975567	EST	
Note: this is a NIH_MGC Library."	BF792513	BF792513	GI:120975567	EST	

Alignment Scores:

Pred. No.:	6,73e-52	Length:	726
Score:	748.00	Matches:	146
Percent Similarity:	94.84%	Conservative:	1
Best Local Similarity:	94.19%	Mismatches:	5
Query Match:	88.21%	Indels:	4
DB:	2	Gaps:	0

US-09-977-418-8 (1-152) x BF792513 (1-726)

QY 1 MetProHisLeuYrYlLeaSPGLyAlPheProIleGlnLeuValArgGluLysLeuHis 20
 Db 151 ATGCCACTCTGTATATGATGGGGTTTTCATACAGCTGCTCTGTATTAACATGCAT 210

QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTTPGly 40
 Db 211 GAAATCCTCGCCGCTCGCCCTGCTGCGGCTCCAGCAAGCCACGCGGGTGGGTG 270

QY 41 TrpGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
 Db 271 TGGGGCTGGTCTCTCTCCCTCCACAGGCTGTCTTCTGGGGCTGCTCCCATGCACACA 330

QY 61 GlySerProAlaArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTTP 80
 Db 331 GGATCACCCTAACAGAGATGAGAGCCA-GGCATGATGGGCTTTGGGTCTCCGAGGTGG 389

QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSer 100
 Db 390 ACCCAGCTCTTCCACCTCCCTCCCTCCGAGCTACAGCTCCATCCATCCCTCTTT 449

QY 100 uileTyGluSerIleGlySerValCysValThr-ThrHisProTyraCysProSer 119
 Db 450 AATATGAAATCTATAGGCTCGCTGTGTATACACACACACCCCTATCTGTCTCTCA 509

QY 120 AsnThrGlnHisTyrHisTTPLeuArgProAsnSerGluLeuSerGlnIleArgPheThr 139
 Db 510 AATCTCAGCATTCACATGAGTGTAGGCAAAATTCACAGCTTCTCTCAATCAGATTACA 569

QY 140 IleSerIlePheIleAsnGlyGluThrSerProSerHis 152
 Db 570 ATCTCCATTTTCATTAAAGGGGAAACATCCCGAGCCAC 608

RESULT 7

W61026 624 bp mRNA linear EST 11-OCT-1996
 LOCUS zc54c09.1 Soares senescent fibroblasts_NBHSF Homo sapiens cDNA
 DEFINITION clone IMAGE:326128 5', mRNA sequence.

W61026 61:1367823
 EST.

ORGANISM Homo sapiens (human)
 SOURCE Homo sapiens

REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 624)
 Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maitra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.

TITLE The Washu-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK

Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNC, contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 982 Std Error: 0.00
 Seq primer: ETPRimer
 High quality sequence stop: 447.

FEATURES

source

Location/Qualifiers

1..624
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="Gene:1257641"
 /db_xref="taxon:9606"
 /clone="IMAGE:326128"
 /tissue_type="senescent fibroblast"
 /lab_host="DH108 (ampicillin resistant)"
 /clone_lib="Soares senescent fibroblasts_NBHSF"
 /note="Vector: pT7T3 (Pharmacia) with a modified
 polylinker V type; phagemid; Site 1: Not I; Site 2: Eco
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)
 primer 15',
 TTTTACCAATCTGAAGTGGAGCGGCCCATTTTCTTTTCTTTTCTTTT 3',
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo."

ORIGIN

Alignment Scores:

Pred. No.:	1.41e-50	Length:	624
Score:	731.00 <td>Matches:</td> <td>134</td>	Matches:	134
Percent Similarity:	97.83% <td>Conservative:</td> <td>1</td>	Conservative:	1
Best Local Similarity:	97.10% <td>Mismatches:</td> <td>2</td>	Mismatches:	2
Query Match:	86.20% <td>Indels:</td> <td>0</td>	Indels:	0
DB:	7	Gaps:	0

US-09-977-418-8 (1-152) x W61026 (1-624)

QY 17 GlnLysLeuHisGluThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLys 36
 Db 1 GATTAACATGATGAATCTCTCCCTCCCTCGCTGGGCTCCACAGCAAGCAAG 60

QY 37 TrpGlyTTPGlyTTPGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAla 56
 Db 61 TGGGGTGGGGGTGGGGGCTGCTCTCTCCCTCCACAGGCTGTCTTGGGGCTGT 120

QY 57 ProMetGlnThrGlySerProAlaArgAspGlySerGlnGlyMetAspGlyAlaLeuGly 76
 Db 121 CCCATGACACAGAGATCACCTAACAGAGATGAGACCGGAGATGATGGGCTTGGGT 180

QY 77 ProArgGlyTTPThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisPr 96
 Db 181 CCTCAGGTTGACCCCGAGCTTCTTCCACCTTCCCTCCGAGCTACGCTCTCATCC 240

QY 96 oSerProSerLeuIleTyGluSerIleGlySerValCysValThr-ThrHisProTyra 116
 Db 241 ATCCCTCTTAACTATGATCTATAGGCTCGGTGTGTATACACACACACCCCTATC 300

QY 116 rGySerProSerAsnThrGlnHisTyrHisTTPLeuArgProAsnSerGluLeuSerGlnI 136
 Db 301 GTTGCTCTCAATACTCGCATTCATTCATGATGTTGGGCAAAATTCAGAGCTTCTCACA 360

QY 136 lAsrPheThrIleSerIlePheIleAsnGlyGluThrSerProSerHis 152
 Db 361 TCAGATTTCATTCATTCATTCATTAACGGGGAACATCCCGAGCCAC 410

RESULT 8

BF306561 636 bp mRNA linear EST 21-NOV-2000
 LOCUS 601888986F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4122788 5',
 DEFINITION mRNA sequence.

W61026 61:1125363
 EST.

ORGANISM Homo sapiens (human)
 SOURCE Homo sapiens

REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 636)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: image.lnln.gov
Plate: LLCM1006 row: 1 column: 21
High quality sequence stop: 634.
Location/Qualifiers

FEATURES
source

1. 636
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:412278"
/tissue_type="thymus/lymphoid"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 17"
/note="Organ: muscle; Vector: pOT7; Site: 1: EcoRI; Site: 2: XhoI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Size-selected >50bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:

Pred. No.: 9,51e-49 Length: 636
Score: 709.00 Matches: 131
Percent Similarity: 97.06% Conservative: 1
Best Local Similarity: 96.32% Mismatches: 2
Query Match: 83.61% Indels: 2
DB: Gaps: 0

US-09-977-418-8 (1-152) x BF306561 (1-636)

QY 1 MetProHisLeuTyrTleAspGlyValPheProIleGlnLeuValArgGluLeuHis 20
DB 230 ATGCCACATCTGATATGATGGGTTTTCATACACCTGCTGATGATACCTGCAT 289
QY 21 GlnThrProAlaValLeuArgLeuGluGlyProProGlyValAlaLeuGlyTyrGly 40
DB 290 GAAACTCCTGCGCTGCTGCGCTGCGGCTCCAGGCAAGCCAAAGGGGTTGGGG 349
QY 41 TrrpGlyTrrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 350 TGGGGCTGCTCTCTCCCTCCACAGGCTGTGTCTTGGGGCTGCTCCCATGACACA 409
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyr 80
DB 410 GGATCACCCTAACAGATGGAAGCCA-GGCATGATGGGGCTTGGGCTCTGAGCTGG 468
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 469 ACCCGAGCTCTTGGCACCTCCCTCCGCGAGCAGCTCCATCCATCCCTCTTT 528
QY 100 uleleYrGlnSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAs 120
DB 529 AATCTATGATCTATAGCTCGCTGCTGTATACACACACACCCCTATCTGCTCTTCAA 588
QY 120 nThrglnHisTyrHisTrrpLeuArgProAsnSerGlnLeuSerGln 135
DB 589 TACTGACATTAACATGCTGAGGCCAAATTCAGAGCTTTCTCAA 634

RESULT 9
AJ707312/c

LOCUS AJ707312 496 bp mRNA linear EST 30-JUN-2004
DEFINITION AJ707312 CPM01 Homo sapiens cdna clone CPM01259, mRNA sequence.
ACCESSION AJ707312
VERSION AJ707312.1 GI:49501607
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 496)

AUTHORS Mallino C.
TITLE Expression profiling of human hypertrophic cardiomyopathy progressed to dilated cardiomyopathy by DNA microarray analysis

JOURNAL

COMMENT Unpublished (2004)
Contact: Mallino C
Biology and CRIBI
University of Padova
Via U. Bassi, 58/B, 35131, ITALY.
Location/Qualifiers

FEATURES
source

1. 496
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CPM01259"
/tissue_type="heart"
/clone_lib="CPM01"
/note="caucasian"

ORIGIN

Alignment Scores:

Pred. No.: 9.69e-47 Length: 496
Score: 683.00 Matches: 128
Percent Similarity: 96.24% Conservative: 0
Best Local Similarity: 96.24% Mismatches: 2
Query Match: 80.54% Indels: 3
DB: Gaps: 0

US-09-977-418-8 (1-152) x AJ707312 (1-496)

QY 23 ProAlaValLeuArgLeuGluGlyProProGlyValAlaLeuGlyTyrGlyTyrGly 42
DB 486 CTTGCCCTCTGCTGCTGCTGCGGCTCCAGGCAAGCCACGCGGGTGGGGCTGGGC 427
QY 43 TrrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThrGlySer 62
DB 426 TGGTCTCTCTCTCCACAGGCTGTGTCTTGGGGCTGCTCCCATGACAGATCA 367
QY 63 ProAsnArgAspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyrThrPro 82
DB 366 CTTAACAGATGGAAGCCAGGCGATGATGGGGCTTGGGCTCTGAGGTTGACCCCA 307
QY 83 AlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLeuTyr 102
DB 306 GCTTCTGCGACCTTCCCTCCGCGAGCTGACCTCTCCATCCATCCCTCTTTATCTA 247
QY 102 rGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerAsnThr 121
DB 246 TGATCTATAGCTGCTGCTGTATACACACACACCCCTATGCTGCTCAATACT 187
QY 122 GlnHisTyrHisTrrpLeuArgProAsnSerGlnLeuSerGlnLeuPheThrIleSer 141
DB 186 CACATTAACATGCTGAGGCCAAATTCAGAGCTTTCAATCAATTCATTC 127
QY 142 IlePheIleAsnGlyGlnThrSerProSerHis 152
DB 126 ATTTTCACTAACCGGGAACATCCCGAGCCAC 94

RESULT 10
CB267751 376 bp mRNA linear EST 20-FEB-2003
LOCUS CB267751
DEFINITION 100657 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
CDNA 5', mRNA sequence.
ACCESSION CB267751

VERSION CB267751.1 GI:28442336
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 376)
 Yang, R.-Z., Shuldiner, A. and Gong, D.-W.
 EST analysis of human adipose gene expression
 Unpublished (2002)

AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Contact: Gong Da-Wei
 Division of Endocrinology, Diabetes and Nutrition
 University of Maryland
 660 Redwood St, HH497, Baltimore, MD 21201, USA
 Tel: 410 706 1672
 Fax: 410 706 1622
 Email: dgong@medicine.umaryland.edu

PCR primers
 FORWARD: CTCGGGAAGCGCCCATGTGTGTG
 BACKWARD: AATGCACTGACTATAGGCGGATTTGG
 Seq primer: GTTGTACCCGGGAATTC.

FEATURES
 source
 1..376
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /sex="Male and Female"
 /tissue_type="Adipose"
 /clone_id="Human Fat Cell 5'-Stretch Plus cDNA library"
 /note="Vector: lambda triplex"

ORIGIN
 Alignment Scores:
 Pred. No.: 1,71e-45 Length: 376
 Score: 666.00 Matches: 122
 Percent Similarity: 97.62% Conservative: 1
 Best Local Similarity: 96.83% Mismatches: 2
 Query Match: 78.54% Indels: 2
 DB: Gaps: 0

US-09-977-418-8 (1-152) x CB267751 (1-376)

QY 13 GlnleuValArgGluLysLeuHisGlnThrProAlaValLeuArgLeuGlyProPro 32
 Db 3 CAGCTGGTGGTGAATAAAGTCAAGAACTCTCCGCTCGCGGCTGCGGCGCTCCA 62

QY 33 GlyLysAlaLysTrpGlyTrpGlyTrpSerPheSerLeuProGlnAlaCysVal 52
 Db 63 GGCAGAGCCAGAGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGT 122

QY 53 **GlyAlaAlaProMetGlnThrGlySerProAsnArgAspGlySerGlnGlyMetAsp 72
 Db 123 CTGGGGGCTGCTCCATGACAGACAGATCACTAACAGAGATGGAAGCCAGGGCATGAT 182

QY 73 GlyAlaLeuGlyProArgGlyTrpThrProAlaSerCysHisLeuProLeuArg-Glnse 92
 Db 183 GGGGCTTGGGTCCTCAGAGGTGGAGCCCAAGCTTCTTGCAACCTTCCCTCCGGGCGAGTC 242

QY 92 AlaLeuHisProSerProSerLeuLeuIleTyrgLysSerIleGlySerValCysValThr-T 112
 Db 243 AGCTTCATCCATCCATCCCTCTTAATCTAAGATCTAAGAGCTCGGTGGTGAACACA 302

QY 112 hrHisProTyArgCysProSerAsnThrGlnHisIleTyrgLysSerValCysValThr 132
 Db 303 CACACCCCTATCGTTGCTTCAATATCAATGATGATGATGATGATGATGATGATGATGAT 362

QY 132 LuLeuSerGlnIle 136
 Db 363 AGCTTTTCAATC 376

RESULT 11
 AJ711705 528 bp mRNA linear EST 30-JUN-2004
 LOCUS

DEFINITION AJ711705 CMPD01 Homo sapiens cDNA clone CMPD09581, mRNA sequence.
 ACCESSION AJ711705
 VERSION AJ711705.1 GI:49506000
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 528)
 Millino, C.
 Expression profiling of human hypertrophic cardiomyopathy
 progressed to dilated cardiomyopathy by DNA microarray analysis
 Unpublished (2004)

AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Contact: Millino C
 Biology and CRIBI
 University of Padova
 Via U. Bassi, 58/B, 35131, ITALY.

FEATURES
 source
 1..528
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CMPD09581"
 /tissue_type="heart"
 /clone_id="CMPD01"
 /note="Caucasian"

ORIGIN
 Alignment Scores:
 Pred. No.: 5.74e-45 Length: 528
 Score: 662.00 Matches: 121
 Percent Similarity: 96.83% Conservative: 1
 Best Local Similarity: 96.03% Mismatches: 2
 Query Match: 78.07% Indels: 2
 DB: Gaps: 0

US-09-977-418-8 (1-152) x AJ711705 (1-528)

QY 29 LeuGlyProProGlyLysAlaLysTrpGlyTrpGlyTrpSerPheSerLeuPro 48
 Db 22 CTGGGGCTCCAGAGGCGCAAGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGTGGGGT 81

QY 49 GlnAlaCysVal**GlyAlaAlaProMetGlnThrGlySerProAsnArgAspGlySer 68
 Db 82 CAGGCTGTGTTCTTGGGGCTGCTCCATGACAGAGATCACTAACAGAGATGGAAGC 141

QY 69 GlnGlyMetAspGlyAlaLeuGlyProArgGlyTrpThrProAlaSerCysHisLeuPro 88
 Db 142 CAGGGCATGATGGGGCTTGGGTCCTCGAGGTTGGAAGCCCAAGCTTCTTGCCACCTTCCC 201

QY 89 LeuArg-GlnSerAlaLeuHisProSerProSerLeuLeuIleTyrgLysSerValGlySerVal 108
 Db 202 CTCCGGGCGAGTGAAGCTTCCATCCATCCCTCTTAATCTAAGATCTAAGAGCTCGGT 261

QY 108 lcyValThr-ThrHisProTyArgCysProSerAsnThrGlnHisIleTyrgLysSerVal 128
 Db 262 GTGTGAACACACACACACCCCTATCGTTGCTTCAATATCAATGATGATGATGATGATGAT 321

QY 128 rGProAsnSerLeuLeuSerGlnIleArgPheThrIleSerIlePheIleAsnGlyGlnT 148
 Db 322 GGCCAATTCAGAGCTTCTCAATATCAATGATGATGATGATGATGATGATGATGATGAT 381

QY 148 hrSerProSerHis 152
 Db 382 CATCCCGAGGCAC 395

RESULT 12
 BF203383 956 bp mRNA linear EST 06-NOV-2000
 LOCUS
 DEFINITION 60186595F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4096646 5',
 mRNA sequence.
 ACCESSION BF203383
 VERSION BF203383.1 GI:11096969

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLNL at: image.llnl.gov
Plate: LHCW65 row: n column: 23
High quality sequence stop: 682.

FEATURES
source
1.9956
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:409846"
/issue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_17"
/note="Organ: muscle; Vector: pOTB7, Site:1: EcoRI;
Site:2: XhoI; CDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:

Pred. No.:	1,256-44	Length:	956
Score:	662.00	Matches:	131
Percent Similarity:	91.78%	Conservative:	3
Best Local Similarity:	89.73%	Mismatches:	12
Query Match:	78.07%	Indels:	3
DB:	2	Gaps:	0

US-09-977-418-8 (1-152) x BF203383 (1-956)

QY 1 MetProHisLeuTyrTlleAspGlyValPheProIleGlnLeuValArgGluLysLeuHis 20
DB 225 ATGCCACATCTGTATATAGATGGGCTTTTCCAAATACAGCTGTTGATTAACATGCAT 284
QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyLYValAlaLysTyrGly 40
DB 285 GAAACTCTGCGCTCTGCGCTGCGCTGCGGCTCCAGGAAAGCCAGCGGGGTTGGGG 344
QY 41 TrrGlyTrrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 345 TGGGGCTGGTCTTCTCCCTCCACAGGCTGTGTTGGGGCTGCTCCATGCAGACA 404
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetLaspGlyValAlaLeuGlyProArgGlyTrr 80
DB 405 GATACCTTAACAGCATGAAAGCCA-GGCATGAGATGGGCTTTGGTCTTGAGGTTGG 463
QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
DB 464 ACCCGAGCTTCTTGCCACCTCCCTCCGGGAGAGCAGTCTCCATCCATCCCTCTTTA 523
QY 101 IleTyrGlnSerIleGlySerValCysValIrrThrHisProTyrArgCysProSerGln 120
DB 524 ATCTAAGATCTAAGGCTGGTGTGTAAACAACACCC-TATCGTGTGCTTCAAT 582
QY 121 ThrGlnHisTyrHisTrrPleuArgProAsnSerGluLeuSerGlnIleArgPheThrIle 140

DB 583 ACTCAGATTA-CATTGGTTGAGGCCAAATTCAGAGCTTTCACATCAGATTAAATC 641
QY 141 SerIlePheIleAsnGly 146
DB 642 TCCATTAAATTAACGGGA 659

RESULT 13
BF203719 976 bp mRNA linear EST 06-NOV-2000
LOCUS 601866380F1 NIH_MGC_17 Homo sapiens CDNA clone IMAGE:4099458 5',
DEFINITION mRNA sequence.
BF203719
ACCESSION BF203719.1 GI:11097305
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLNL at: image.llnl.gov
Plate: LHCW67 row: p column: 19
High quality sequence stop: 676.

FEATURES

source

1.976
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4099458"
/issue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_17"
/note="Organ: muscle; Vector: pOTB7, Site:1: EcoRI;
Site:2: XhoI; CDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:

Pred. No.:	1,296-44	Length:	976
Score:	662.00	Matches:	137
Percent Similarity:	90.85%	Conservative:	2
Best Local Similarity:	89.54%	Mismatches:	13
Query Match:	78.07%	Indels:	6
DB:	2	Gaps:	0

US-09-977-418-8 (1-152) x BF203719 (1-976)

QY 1 MetProHisLeuTyrTlleAspGlyValPheProIleGlnLeuValArgGluLysLeuHis 20
DB 229 ATGCCACATCTGTATATAGATGGGCTTTTCCAAATACAGCTGTTGATTAACATGCAT 288
QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyLYValAlaLysTyrGly 40
DB 289 GAAACTCTGCGCTCTGCGCTGCGGCTCCAGGAAAGCCAGCGGGTGGGG 348
QY 41 TrrGlyTrrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 349 TGGGGCTGGTCTTCTCCCTCCACAGGCTGTGTTGGGGCTGCTCCATGCAGACA 408

QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
 DB 409 GGATCACCTTAACAAGATGGAAGCA-GGCATGATGGGGC-TTGGGTCCTCGAGTTGG 466
 QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 DB 467 ACCCAGCTTCTTGCCACCTTCCCTCCGGGCAATCAGCTCTCCATCCATCCCTCTTT 526
 QY 100 uUeTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerAs 120
 DB 527 AATCTATGAATCTATAGGCTCGGTGTGTGTACACACACCC-TATCGTTGTGCTTCAA 585
 QY 120 nThrGlnHisTyrHisThrLeuArgProAsnSerGlnLeuSerGlnIleArgPheTrrI 140
 DB 586 TACTAGCATTAACATGATGGTGTGAGGCCAAT-TCAAGACT-TCTCAAAATCAGATTTCACAT 643
 QY 140 eSerIlePheIleAsnGlyGlnThrSerProSerHis 152
 DB 644 CTCATTCTCATTAACGGGGACAGATCCCGAGCCAT 680

RESULT 14

LOCUS CD625679 518 bp mRNA linear EST 12-JAN-2004
 DEFINITION 5605057601 FLP Homo sapiens cDNA, mRNA sequence.
 ACCESSION CD625679
 VERSION CD625679.1 GI:40273945
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE (bases 1 to 518)
 JOURNAL Circular rapid amplification of cDNA ends for high-throughput
 COMMENT extension cloning of partial genes
 Genomics 84 (1), 205-210 (2004)
 CONTACT: Fu GK
 INCYTE Genomics, Inc.
 3160 Porter Dr., Palo Alto, CA 94304, USA
 Tel: 6508454102
 Email: gfu@incyte.com

FEATURES
 source Location/Qualifiers
 1..518
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="FLP"
 /note="Vector: pDrive Cloning Vector"

ORIGIN
 Alignment Scores:
 Pred. No.: 2,05e-42 Length: 518
 Score: 631.00 Matches: 117
 Percent Similarity: 95.93% Conservative: 1
 Best Local Similarity: 95.11% Mismatches: 3
 Query Match: 74.41% Indels: 2
 DB: 6 Gaps: 0

US-09-977-418-8 (1-152) x CD625679 (1-518)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlyLeuHis 20
 DB 145 ATGCCACATCTGATATAGATGGGGTTTTCACATACAGCTGCTCGATTAACATGCAT 204
 QY 21 GlnThrProAlaValLeuArgLeuGlyProProGlyLysAlaLysTyrGlyTyrGly 40
 DB 205 GAACCTCCTGCGCTGCTGCGCTGCTGCGGCTCCAGGCAAGCCACGTGGGGTGGGG 264
 QY 41 TrrGlyTrrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnTrr 60
 DB 265 TGGGGTGGTCTCTCTCCCTCCACAGGCGCTGTCTTGGGGCTGCTCCCATGACAGACA 324

QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
 DB 325 GGATCACCTTAACAAGATGGAAGCA-GGCATGATGGGGC-TTGGGTCCTCGAGTTGG 384
 QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 DB 385 ACCCAGCTTCTTGCCACCTTCCCTCCGGGCAATCAGCTCTCCATCCATCCCTCTTT 444
 QY 100 uUeTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerAs 120
 DB 445 AATCTATGAATCTATAGGCTCGGTGTGTGTACACACACCC-TATCGTTGTGCTTCAA 504
 QY 120 snThr 121
 DB 505 ATRACT 509

RESULT 15

LOCUS B1113724 562 bp mRNA linear EST 26-JUN-2001
 DEFINITION 602860919P1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:5020037 5',
 mRNA sequence.
 ACCESSION B1113724
 VERSION B1113724.1 GI:14564625
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NIH-MGC http://mgi.mol.nh.gov/.
 TITLE (bases 1 to 562)
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: ggapbs-remail.nih.gov

FEATURES
 source Location/Qualifiers
 1..562
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:5020037"
 /tissue_type="Thadomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 17"
 /note="Organ: muscle; Vector: pOT7; Site 1: EcoRI;
 Site 2: XhoI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:
 Pred. No.: 2.13e-38 Length: 562
 Score: 583.00 Matches: 107
 Percent Similarity: 96.43% Conservative: 1
 Best Local Similarity: 95.54% Mismatches: 3
 Query Match: 68.75% Indels: 1
 DB: 4 Gaps: 0

US-09-977-418-8 (1-152) x B1113724 (1-562)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlyLeuHis 20

```

Db      226 ATGCCACATCTGTATATAGATGGGTTTTTCCATACAGCTGTCGTGATAAAGTGCAT 285
Qy      21  GIUTHPROALAValleuArgLeuGlyProProGlyLysAlaLysTrpGly 40
Db      286 GAAACTCTGCGCGCTGCTGCGCGCTGCGCGCTGCCAGCGCAGCGCGCGCTGCGCGG 345
Qy      41  TrpGlyTrpSerPheSerLeuPProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
Db      346 TGGGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 405
Qy      61  GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTrp 80
Db      406 GGATCACCTTACAGAGATGAAAGCAGAGGCGATGATGGGCTTAGGGTCTCGAGGTGG 465
Qy      81  ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
Db      466 ACCCCAGCTTCTTGCCACCTTCCCTCCGCGCAGTCAGCTCTCCATCCATCCCTCTTT 525
Qy      100 uIeTyrgIuSerIeGlySerValCysValThr 111
Db      526 AATCTATGAATCTATAGGCTCGGTGTGTACA 559

```

Search completed: October 24, 2004, 17:57:19
 Job time : 2710 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 24, 2004, 16:03:27 ; Search time 3547 Seconds
(without alignments)
2026.511 Million cell updates/sec

Title: US-09-977-418-8
Perfect score: 848
Sequence: 1 MPELVYGVFPIQLVREKJH.....LSQIRPTISIFINGETSPSH 152

Scoring table:
BLOSUM62
Xgapod 10.0 , Xgapext 0.5
Ygapod 10.0 , Ygapext 0.5
Fgapod 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4526729 seqs, 23644849745 residues
Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgm2_1/USPTO.spool/US09977418/runatc.22102004.115947.1017/app.query.fasta_1.327
-DB=GenEmbl -OPMT=fastap -SUFFIX=gcg -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ps -NOR=gcg -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09977418.@CGN_1.1.3731.@runatc.22102004.115947.1017 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-FEA_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOD=10 -XGAPEXT=0.5 -FGAPOD=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : (GenEmbl):
1: gb_ha:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	846	99.8	1987	6	AX058600 Sequence
2	817	96.3	2152	9	AF155908 Homo sapi
3	817	96.3	2152	9	HS0243191 Homo sapi
4	817	96.3	2747	9	HS0803468 Homo sapi

Result No.	Score	Query Match	Length	DB ID	Description
5	817	96.3	109377	9	AL355994 Human DNA
6	811	95.6	2138	6	CO721116 Sequence
7	811	95.6	2192	6	AX017502 Sequence
8	811	95.6	2192	6	BD135192 Human nuc
9	811	95.6	2589	9	BC006319 Homo sapi
10	811	95.6	2701	6	CO783843 Sequence
11	811	95.6	2701	6	BD127861 Primer fo
12	811	95.6	2701	9	AK075151 Homo sapi
13	807	95.2	2142	9	AK095911 Homo sapi
14	807	95.2	76241	9	HS013364 Human DNA
15	802	94.6	2410	6	AX747447 Sequence
16	802	94.6	2410	9	AK092275 Homo sapi
17	797	94.0	2163	9	BSX53790 Homo sapi
18	772.5	91.1	601	11	EV167837 sqm6636
19	772.5	91.1	2303	11	EV178365 sqm10009
20	578	68.2	1550	6	AX081428 Sequence
21	512	60.4	305	6	AX058658 Sequence
22	382	42.7	201	11	EV203380
23	233	27.5	306769	2	AL556105
24	230	27.1	750	11	EV166340 HSPB7_990
25	112.5	13.3	188317	2	AC148957 Octleumir
26	112.5	13.2	105628	2	AC138144 Homo sapi
27	111.5	13.1	62295	2	AC135730 Homo sapi
28	111.5	13.1	153152	9	AC021899 Homo sapi
29	111.5	13.1	161612	9	AC020754 Homo sapi
30	111.5	13.1	161629	2	AC087604 Pan trogl
31	111.5	13.1	209168	9	AC090955 Homo sapi
32	111	13.1	180876	9	AC021317 Homo sapi
33	110	13.0	123925	9	AC003976 Homo sapi
34	110	13.0	163628	2	AC139367 Homo sapi
35	110	13.0	177255	2	AC148806 Homo sapi
36	110	13.0	193646	9	AC131056 Homo sapi
37	109.5	12.9	208596	2	AC127589 Mus muscu
38	108.5	12.8	4268	9	HMALPDP M15159 Human place
39	108.5	12.8	143929	10	AC124714 Mus muscu
40	108.5	12.8	165196	10	AL596123 Mouse DNA
41	108.5	12.8	169247	9	AL669970 Human DNA
42	108.5	12.8	343504	2	AL158031 Homo sapi
43	108	12.7	96678	9	AL096755 Homo sapi
44	108	12.7	129989	9	AP006621 Homo sapi
45	108	12.7	178778	9	AC132936 Homo sapi

ALIGNMENTS

RESULT 1	LOCUS	SEQUENCE	7 from Patent WO0075321.	DNA	linear	PAT 17-JAN-2001
AX058600	AX058600	Sequence 7 from Patent WO0075321.				
AX058600	AX058600	Sequence 7 from Patent WO0075321.				
AX058600.1	GI:12310942					
Homo sapiens (human)						
Homo sapiens						
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
CURATOR						
FEATURES						
source						
1..1987						
/organism="Homo sapiens"						
/mol_type="unassigned DNA"						
/db_xref="taxon:9606"						
991..1449						
/note="unamed protein product"						
/codon_start=1						
/protein_id="CAC22515.1"						
/db_xref="gi:12310942"						
/translation="MPELVYGVFPIQLVREKJHETPAVLELGPCKAKKMGKMGWS FSLPQACVXGAHPWQTSGRNDRSGQMDGALGPNGWTPASCHLPLRQSLAPSPSLIY						

ORIGIN

ESIGSVCTTPHPYPCPSNTQHYHWMRPNSELSQIRFTISIFINGETSPSH"

Alignment Scores:

Pred. No.:	7, 27e-53	Length:	1987
Score:	846.00	Matches:	152
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.76%	Indels:	0
DB:	6	Gaps:	0

US-09-977-418-8 (1-152) x AX058600 (1-1987)

QY 1 MetProHisLeuTyrTlleaspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20

DB 991 ATGCCACACCTGTATATAGATGGGGTTTTCATACAGAGCTGGTGTGAAAACTGAT 1050

QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyLysValAlaLysTyrGlyTyrGly 40

DB 1051 GAAATCCCTGCGCTCTGCGCTGCTGCGGCTCCAGCGAGGCCAAGTGGGGTGGGG 1110

QY 41 TrrpGlyTrrpSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThr 60

DB 1111 TGGGGCTGGTCTCTCCCTCCACAGGCTGTTCNTGGGGCTGCTCCATCGAGACA 1170

QY 61 GlySerProAsnArgaspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyr 80

DB 1171 GGATCCCTTAACAGATGGAAGCCAGGCAATGATGGGGCTTTGGGCTCTCGAGGTTGG 1230

QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100

DB 1231 ACCCGAGCTTCTTGCCACCTCCCTCCGCGAGTCAGCTCTCATCCATCCCTCTCTTA 1290

QY 101 IleTyrGlnSerIleGlySerValCysValThrThrHisProTyrArgCysProSerHis 120

DB 1291 ATCTATGAATCTATAGCTCGGTGGTGTGTGAACAACACCCCTATCGTGTTCCTCAAT 1350

QY 121 ThrGlnHisTyrHisTrrpLeuArgProAsnSerGlnLeuSerGlnIleAspPheThrIle 140

DB 1351 ACACGACATTTACCATGTTGTTAGGCCAATTCCAGAGCTTTCATATCGATTTCATATC 1410

QY 141 SerIlePheIleAsnGlyGlnThrSerProSerHis 152

DB 1411 TCCATTTTCATTAAACGGGAAACATCCCGAGCCAC 1446

RESULT 2 AF155908 2152 bp mRNA linear PRI 27-DEC-1999

LOCUS Homo sapiens cardiovascular heart shock protein mRNA, complete cds.

ACCESSION AF155908.1 GI:6635998

VERSION AF155908.1

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 2152)

AUTHORS Krieff,S., Faivre,J.F., Robert,P., Le Douarin,B.,

TITLE Brument-Latignon,N., Lefevre,I., Bouzyk,M.M., Anderson,K.M.,

JOURNAL Identifier and characterization of cvhsp. A novel human small

MEDLINE stress protein selectively expressed in cardiovascular and

PUBMED insulin-sensitive tissues

AUTHORS J. Biol. Chem. 274 (51), 36592-36600 (1999)

REFERENCE 2 (bases 1 to 2152)

TITLE Krieff,S., Faivre,J.F., Robert,P., Le Douarin,B.,

JOURNAL Brument-Latignon,N., Lefevre,I., Bouzyk,M.M., Anderson,K.A.,

MEDLINE Greller,L.D., Tobin,F.L., Souchet,M. and Brill,A.

PUBMED Direct Submission

SUBMITTED (02-JUN-1999) Cardiovascular Pharmacology, SmithKline

Beecham, 4 rue du Chesnay-Beauregard, Saint-Gregoire 35760, France

FEATURES

source

Location/Qualifiers
1..2152
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p36.23-p34.3 between markers DIS434 and DIS507"
75..587
/note="25 kDa heat stress protein; cvhsp, preferentially
expressed in cardiovascular and insulin-sensitive tissues;
interacts with C-terminal domain of alpha-filamin or
actin-binding-280, ABP-280"
/product="cardiovascular heat shock protein"
/protein_id="AA020022.1"
/db_xref="GI:6635998"
/translation="MSHRTSSTFRARSHSSSSSSSSSSSSSSASAPALPAQDPPEKA
LSMFSDPDSFMRPHEPAPFAPRPGANIKITLDPAFAVDVDFSEDIITVTSN
NHIEVAERKLADGVTMTFAHKQLPEDVDEPTSVTSALREDSGLTIRARRPHEHV
QOTFEETIKI"
2155..2120

CDS

polya_signal

2155..2120

ORIGIN

polya_signal

2155..2120

Alignment Scores:

Pred. No.:	1, 04e-50	Length:	2152
Score:	817.00	Matches:	150
Percent Similarity:	98.05%	Conservative:	1
Best Local Similarity:	97.40%	Mismatches:	2
Query Match:	96.34%	Indels:	0
DB:	9	Gaps:	0

US-09-977-418-8 (1-152) x AF155908 (1-2152)

QY 1 MetProHisLeuTyrTlleaspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20

DB 1201 ATGCCACATCTGTATATAGATGGGGTTTTCATACAGCTGGTGTGATAAAGCAT 1260

QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyLysValAlaLysTyrGlyTyrGly 40

DB 1261 GAAATCCCTGCGCTCTGCGCTGCGGCGCTCCAGGCAAGGCCAAGGGGGTGGGG 1320

QY 41 TrrpGlyTrrpSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThr 60

DB 1321 TGGGGCTGGTCTCTCCCTCCACAGGCTGTGTTCTTGGGGCTCTCCATCGAGACA 1380

QY 61 GlySerProAsnArgaspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyr 80

DB 1381 GGATCACCTTAACAGATGGAAGCCAGGCGATGATGGGGCTTGGGCTCTCGAGGTTGG 1440

QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLeu 100

DB 1441 ACCCGAGCTTCTTGCCACCTCCCTCCGCGAGTCAGCTCTCATCCATCCCTCTCTTA 1500

QY 100 uileTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA 120

DB 1501 AATCTATGAATCTATAGCTCGGTGGTGTGTGAACAACACCCCTATCGTGTCTTCA 1560

QY 120 snThrGlnHisTyrHisTrrpLeuArgProAsnSerGlnLeuSerGlnIleAspPheThrI 140

DB 1561 ATACTCGACATTCATTCATGTTGAGGCCAATTCACAGGCTTCTCAATACAGATTACAA 1620

QY 140 leSerIlePheIleAsnGlyGlnThrSerProSerHis 152

DB 1621 TCTCATTTTCATTAAACGGGAAACATCCCGAGCCAC 1656

RESULT 3 HSA243191 2152 bp mRNA linear PRI 14-DEC-1999

LOCUS Homo sapiens mRNA for cardiovascular heart shock protein.

ACCESSION AJ243191.1 GI:6589425

VERSION AJ243191.1

KEYWORDS cvhsp gene; heat shock protein.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
 Krief,S., Faivre,J.F., Robert,P., le Douarin,B.,
 Brument-Larignon,N., Lefriere,I., Bouzyk,M.M., Anderson,K.M.,
 Gellier,L.D., Tobin,F.L., Souchet,M. and Bril,A.
 Identification and characterization of cvhsp, A novel human small
 stress protein selectively expressed in cardiovascular and
 insulin-sensitive tissues

JOURNAL J. Biol. Chem. 274 (51), 36592-36600 (1999)

MEDLINE 20062883
 PUBMED 10593960
 2 (bases 1 to 2152)

REFERENCE Krief,S.
 AUTHORS Direct Submission
 TITL Submitted (17-JUN-1999) Krief S., Cardiovascular Pharmacology,
 JOURNAL Smlthline Beecham, 4 rue du Chesnay Beauregard - Saint-Gregoire,
 35760, FRANCE

FEATURES
 source location/Qualifiers
 1. 2152
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /tissue_type="heart"
 /dev_stage="fetal"
 1. 2152
 /gene="cvhsp"
 75..587
 /gene="cvhsp"
 /function="small heat stress protein"
 /codon_start=1
 /product="heat shock protein"
 /protein_id="CA863258.1"
 /db_xref="GI:6580426"
 /db_xref="GOA:Q9UB93"
 /db_xref="Swiss-Prot:Q9UB93"
 /translation="MSHRTSSTFRAERSFHSSSSSSSTSSASRALPAQDPMEKA
 LMSFSDPFGFMRPSEPLPARPGAGNITLGDAYEFADVDFSPEDIVITSN
 NHIEVRAEKLAADGVTAVMTFAHKQCLPEVDPTSVMSALREDSLTIRARHPHTSHV
 CQFRTREIKT"
 polyA_signal 2115..2120
 /gene="cvhsp"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.04e-50 Length: 2152
 Score: 817.00 Matches: 150
 Percent Similarity: 98.05% Conservative: 1
 Best Local Similarity: 97.40% Mismatches: 2
 Query Match: 96.34% Indels: 0
 DB: Gaps: 0

US-09-977-418-8 (1-152) x HSA243191 (1-2152)

QY 1 MetProHsLeuTYrTlleaSPGLyAlPheProIleGlnLeuValArgGluYsLeuHs 20
 Db 1201 ATGCCAATCTGTATATAGATGGGGTTTTCCTCAATACAGCTGCTGTATTAACATGCAT 1260

QY 21 GluThrProAlaValleuArgLeuLeuGlyPProProGlyLySaAlaLysTPGlyTTPGly 40
 Db 1261 GAAACTCTGCGCTGCTGCGCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGG 1220

QY 41 TTPGLYTPSePReSeRLeuPProGlnAlaCyVal**GlyAlaAlaPProMetGlnThr 60
 Db 1221 TGGGGCTGGTCTCTCTCCCTCCACAGCCCTGTGTTTGGGGCTGCTCCACATGCACACA 1280

QY 61 GlySePProAnaGyASPGLySerGlnGlyWheAspGlyAlaLeuGlyPProAspGlyTTP 80
 Db 1281 GGATCCTTAACAGATGAGACCGCAGGCGATGATGGGGCTTTGGGTCTCTCAAGTTGG 1440

QY 81 ThrProAlaSerCySHsLeuProLeuArg-GlnSerAlaLeuHsPProSerProSerLe 100

Db 1441 ACCCAGCTTCTTGGCACCCTTCCCTCCGGGAGAGACACTCTCCATCCATCCCTCTTT 1500

QY 100 vtleTygIuseRlleGlySerValCyValThr-ThHisProTyArgCySPoseRa 120
 Db 1501 AATCTATGATCTATATAGCTCGTGTGTGTATACACACACACCCCTATGTTGCTTCA 1560

QY 120 snThGlnHsTyHsThRpleuArgProAnseRgLeuSerGlnLeuArgPheThRi 140
 Db 1561 ATACTGACATTCATTCAGCTGTTGAGCCCAATTCAGACTTCTTCAATCATTTACAA 1620

QY 140 lseRllePheHleasnglyGlnThrSerProSerHs 152
 Db 1621 TCTCATTTTCATTAACGGGAAACATCCCGAGCCAC 1658

RESULT 4
 HSM603488 2747 bp mRNA linear PRI 13-MAY-2003
 LOCUS Homo sapiens mRNA; cDNA DKFZp686P1416 (from clone DKFZp686P1416).
 DEFINITION AL832181
 ACCESSION AL832181
 VERSION GI:21732726
 KEYWORDS

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2747)
 AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Aml,D.C., Osanger,A.,
 Fobo,G., Han,M. and Wiemann,S.
 Direct Submission
 Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764
 Neuherberg, GERMANY

TITLE Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 JOURNAL Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 COMMENT sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 This clone (DKFZp686P1416) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
 information about the clone and the sequencing project is available
 at http://mips.gsf.de/proj/cDNA/.

FEATURES
 source location/Qualifiers
 1. 2747
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZp686P1416"
 /tissue_type="human cervix"
 /clone_lib="686 (synonym: h1cc3). Vector pSPori1_Sfi; host
 DH10B; sites SfiI + SfiIB"
 /dev_stage="adult"
 polyA_signal 2626..2631
 polyA_site 2641

ORIGIN
 Alignment Scores:
 Pred. No.: 1.33e-50 Length: 2747
 Score: 817.00 Matches: 150
 Percent Similarity: 98.05% Conservative: 1
 Best Local Similarity: 97.40% Mismatches: 2
 Query Match: 96.34% Indels: 0
 DB: Gaps: 0

US-09-977-418-8 (1-152) x HSM603488 (1-2747)

QY 1 MetProHsLeuTYrTlleaSPGLyAlPheProIleGlnLeuValArgGluYsLeuHs 20
 Db 1711 ATGCCAATCTGTATATAGATGGGGTTTTCCTCAATACAGCTGCTGTATTAACATGCAT 1770

QY 21 GluThrProAlaValleuArgLeuLeuGlyPProProGlyLySaAlaLysTPGlyTTPGly 40
 Db 1771 GAAACTCTGCGCTGCTGCGCTGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGG 1830

QY 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
 Db 1831 TGGGGGTGGCTCTTCCCTCCACAGAGGCTGTGTCTTGGGGCTGCTCCATGACAGCA 1890
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTTP 80
 Db 1891 GGATTCACCTTACAGAGATGGAACCCAGGCGATGAGAGGCTTGGGCTCTTGAGAGTTGG 1950
 QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 Db 1951 ACCCGAGCTCTTGGCACCTTCCCTCCGGGAGTACAGCTCCATCCATCCCTCTCTT 2010
 QY 100 uileTyrgLuserlileGlySerValCysValThr-ThrHisProTyArgCysProSerA 120
 Db 2011 AATCTTGAATCTTATGCTCGCTGCTGTGTAACACACACACCCCTTATCTTCTTCA 2070
 QY 120 snThrglnHisTyriHisTTPLeuArgProAsnSerGluLeuSerGlnIleArgPheThr 140
 Db 2071 ATACTCAGCATTCACATGCTGTGAGGCCAAATTCAGAGCTTCTCAATCAGATTACAA 2130
 QY 140 leSerllePheIleAsnGlyGluThrSerProSerHis 152
 Db 2131 TCTCCATTTCATTAAAGGAGAAACATCCCGAGCCAC 2168
 RESULT 5
 AL355994/c 109377 bp DNA linear PRI 10-JUN-2003
 LOCUS Human DNA sequence from clone RP11-5P18 on chromosome 1, complete
 DEFINITION sequence.
 ACCESSION AL355994 GI:31616669
 VERSION AL355994.11
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 109377)
 DIRECT SUBMISSION
 Submitted (10-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Jun 11, 2003 this sequence version replaced gi:31616669.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone configs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Ch1> RP11-5P18 is from the library RP11-11.1 constructed by the group of

Pieter de Jong. For further details see <http://www.choi1.org/bacpac/home.htm>
 VECTOR: pBAC3.6.
 location/Qualifiers
 1..109377
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-5P18"
 /clone_1fb="RP11-11.1"

ORIGIN

Alignment Scores:
 Pred. No.: 5.55e-49 Length: 109377
 Score: 817.00 Matches: 150
 Percent Similarity: 98.05% Conservative: 1
 Best Local Similarity: 97.40% Mismatches: 1
 Query Match: 96.34% Indels: 2
 DB: Gaps: 0

US-09-977-418-8 (1-152) x AL355994 (1-109377)

QY 1 MetProHisLeuTyriIleAspGlyValPheProIleGlnLeuValArgGlyLeuHis 20
 Db 46193 ATGCCACATCTGTAATAGATGGGGTTTTCATACAGCTGCTGTGATAACTGCAAT 46134
 QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyValAlaValTTPGlyTTPGly 40
 Db 46133 GAAATCTCTCCGCTCCGCTGCTGGGCTCTCCAGGCAAGCCCAAGAGGGGTGGG 46074
 QY 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
 Db 46073 TGGGGGTGGCTCTTCCCTCCACAGAGGCTGTGTTCTTGGGGCTGCTCCATGACAGCA 46014
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTTP 80
 Db 46013 GGATTCACCTTACAGAGATGGAACCCAGGCGATGAGAGGCTTGGGCTCTTGAGAGTTG 45954
 QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 Db 45953 ACCCGAGCTCTTGGCACCTTCCCTCCGGGAGTACAGCTTCCATCCATCCCTCTT 45894
 QY 100 uileTyrgLuserlileGlySerValCysValThr-ThrHisProTyArgCysProSerA 120
 Db 45893 AATCTTGAATCTTATGCTCGCTGCTGTGTAACACACACACCCCTTATCTTCTTCA 45834
 QY 120 snThrglnHisTyriHisTTPLeuArgProAsnSerGluLeuSerGlnIleArgPheThr 140
 Db 45833 ATACTCAGCATTCACATGCTGTGAGGCCAAATTCAGAGCTTCTCAATCAGATTACAA 45774
 QY 140 leSerllePheIleAsnGlyGluThrSerProSerHis 152
 Db 45773 TCTCCATTTCATTAAAGGAGAAACATCCCGAGCCAC 45736

RESULT 6

LOCUS CO721116 2138 bp DNA linear PAT 03-FEB-2004
 DEFINITION Sequence 7050 from Patent WO2068579.
 ACCESSION CO721116
 VERSION CO721116.1 GI:42281973

SOURCE

ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
 Kites, such as nucleic acid arrays, comprising a majority of
 humanexons or transcripts, for detecting expression and other uses
 thereof
 Patent: WO 02068579-A 7050 06-SEP-2002;
 JOURNAL PE Corporation (NY) (US)


```

FEATURES
  Source
    Location/Qualifiers
      1..2138
        /organism="Homo sapiens"
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"
ORIGIN
Alignment Scores:
  Pred. No.:      2,84e-50      Length:      2138
  Score:          811.00        Matches:     149
  Percent Similarity: 97.40%    Conservative: 1
  Best Local Similarity: 96.75%  Mismatches:  2
  Query Match:    95.64%       Indels:      2
  DB:             Gaps:        0
US-09-977-418-8 (1-152) x CQ221116 (1-2138)
QY      1 MePProHsLeuTYrIleAspGlyValPheProIleGlnLeuValArgGluLysLeuHis 20
Db      1203 ATGCCACATCTGTATATAGATGGGGTTTTCATATACAGCTGGTCTGTATTAATGCAAT 1262
QY      21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTyrGly 40
Db      1263 GAAACTCTCGCCGCTCCGCGCTGCTGGGGCTCCAGGCAAGGCCACGTTGGGGGG 1322
QY      41 TrpGlyTyrPheSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
Db      1323 TGGGGCTGGTCTCTCTCCCTCCACAGCCCTGtGtTCTTGGGGCTGCTCCCATGCGAGACA 1382
QY      61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
Db      1383 GGATCACCCTAACAGATGAGATGAGAGCCAGGGCATGAGTGGGGCTTGGGCTCCGAGGTGG 1442
QY      81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerIle 100
Db      1443 ACCCCACCTCTTGGCCACCTTCCCTCCGGGCACTGCTCCATCCATCCCCCTCTTT 1502
QY      100 uileTyrGluSerIleGlySerValCysValThr-ThrHisProTyrArgCysProserA 120
Db      1503 AATATAGATCTATAGGCTCGGTGTGTATACACACACACCCCTATCGTTGCTTCA 1562
QY      120 snThrGlnHisTyrHisTyrLeuArgProAsnSerGluLeuSerGlnIleArgPheThrI 140
Db      1563 ATACTCGACATTACCATTTGTTGAGGCCAATTCAGAGCTTTCTCAATCAGATTTCACA 1622
QY      140 leSerIlePheIleAsnGlyGluThrSerProSerHis 152
Db      1623 TCTCCATTTCATTAAACGGGAAACATCCCGAGCCAC 1660
RESULT 7
  LOCUS      AX017502      2192 bp      DNA      linear      PAT 07-SEP-2000
  DEFINITION Human nucleic acid sequence from normal breast tissue
  ACCESSION  AX017502
  VERSION    AX017502.1  GI:10042299
  KEYWORDS
  ORGANISM   Homo sapiens (human)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  AUTHORS    Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
  TITLE      Human nucleic acid sequences from normal breast tissue
  JOURNAL    Patent: WO 9947655-A 48 23-SEP-1999;
  SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
  BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
  (DE); PILARSKY CHRISTIAN (DE)
FEATURES
  Source
    Location/Qualifiers
      1..2192
        /organism="Homo sapiens"
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"
ORIGIN
Alignment Scores:
  Pred. No.:      2,91e-50      Length:      2192
  Score:          811.00        Matches:     149
  Percent Similarity: 97.40%    Conservative: 1
  Best Local Similarity: 96.75%  Mismatches:  2
  Query Match:    95.64%       Indels:      2
  DB:             Gaps:        0
US-09-977-418-8 (1-152) x AX017502 (1-2192)
QY      1 MePProHsLeuTYrIleAspGlyValPheProIleGlnLeuValArgGluLysLeuHis 20
Db      1198 ATGCCACATCTGTATATAGATGGGGTTTTCATATACAGCTGGTCTGTATTAATGCAAT 1257
QY      21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTyrGly 40
Db      1258 GAAACTCTCGCCGCTCCGCGCTGCTGGGGCTCCAGGCAAGGCCACGTTGGGGGG 1317
QY      41 TrpGlyTyrPheSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
Db      1318 TGGGGCTGGTCTCTCTCCCTCCACAGCCCTGtGtTCTTGGGGCTGCTCCCATGCGAGCA 1377
QY      61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
Db      1378 GGATCACCCTAACAGATGAGATGAGAGCCAGGGCATGAGTGGGGCTTGGGCTCCGAGGTGG 1437
QY      81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerIle 100
Db      1438 ACCCCACCTCTTGGCCACCTTCCCTCCGGGCACTGCTCCATCCATCCCCCTCTTT 1497
QY      100 uileTyrGluSerIleGlySerValCysValThr-ThrHisProTyrArgCysProserA 120
Db      1498 AATCTATGATCTATAGGCTCGGTGTGTATACACACACCCCTATCGTTGCTTCA 1557
QY      120 snThrGlnHisTyrHisTyrLeuArgProAsnSerGluLeuSerGlnIleArgPheThrI 140
Db      1558 ATACTCAGCATTAACCATTTGTTGAGGCCAATTCAGAGCTTTCTCAATCAGATTTCACA 1617
QY      140 leSerIlePheIleAsnGlyGluThrSerProSerHis 152
Db      1618 TCTCCATTTCATTAAACGGGAAACATCCCGAGCCAC 1655
RESULT 8
  LOCUS      BD135192      2192 bp      DNA      linear      PAT 18-SEP-2002
  DEFINITION Human nucleic acid sequence originating in normal mammary tissue.
  ACCESSION  BD135192
  VERSION    BD135192.1  GI:23230137
  KEYWORDS   JP 2002506639-A/39.
  SOURCE     Homo sapiens (human)
  ORGANISM   Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  1 (bases 1 to 2192)
  Speitt,T., Hinzmann,B., Schmitt,A., PilarSKI,C., Duhl,E. and
  Rosenthal,A.
  TITLE      Human nucleic acid sequence originating in normal mammary tissue
  JOURNAL    Patent: JP 2002506639-A 39 05-MAR-2002;
  METAGEN GESELLSCHAFT FUER GENOME FORSCHUNG MBH
  COMMENT
  OS Homo sapiens (human)
  PN JP 2002506639-A/39
  PD 05-MAR-2002
  PF 19-MAR-1999 JP 2000536838
  PR 20-MAR-1998 DE 198 13 835.0
  PI THOMAS SPEFT, BERND HINTZMAN, ARMIN SCHMITT, CHRISTIAN PIRARSKI,
  PI EDGAR DUHL,
  PI ANDRE ROSENTHAL
  PC C12N15/09,A61K48/00,A61P35/00,A61P43/00,A61P43/00,C07K14/47,
  PC C07K16/18,
  PC C12N1/21,C12N5/10,C12P21/02,C12Q1/68,G01N33/68//A61K38/00, PC
  C12N15/00.

```

```

PC C12N5/00/A61K37/02
CC Human nucleic acid sequence originating in normal mammary CC
FH Key tissue
FT source 1..2192
Location/Qualifiers
  /organism="Homo sapiens (human)"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"

FEATURES
  source
    Location/Qualifiers
      1..2192
        /organism="Homo sapiens (human)"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"

ORIGIN
Alignment Scores:
  Alignment No.: 2,91e-50
  Score: 811.00
  Percent Similarity: 97.40%
  Best Local Similarity: 96.75%
  Query Match: 95.64%
  DB: 6
  Gaps: 0

US-09-977-418-8 (1-152) x BD135192 (1-2192)
QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGluIysLeuHis 20
Db 1198 ATCCCATCTGTATTAATGGGCTTTTCCATACAGCTGTCGATTAACCTGCAT 1257
QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyValAlaAlaIleTyrGly 40
Db 1258 GAACACTCGCGCTCTGCGCTGCGCTGCGCGCTCCGACGACGCGGCGGCTGGGG 1317
QY 41 ThrGlyTyrSerPheSerLeuProGlnIleCysVal**GlyAlaAlaProMetGlnThr 60
Db 1318 TGGGCGTGGCTCTTCCCTCCCAAGGCTGTGTCTTGGGCGTGTCTCCATGACGACA 1377
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
Db 1378 GATACACCTAACAGAAATGAGAGCCAGGACGATGAGGCGCTTGGGCTCTGAGGTTGG 1437
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
Db 1438 ACCCGAGCTTCTTGGCACCTTCCCTCCCGGACGATGAGCTCCATCCATCCCTCTTT 1497
QY 100 uileTyrGlySerIleGlySerValCysValThr-ThrHisProTyrArgCysProserA 120
Db 1498 AATCTATGATCATATGCTCGCTCGCTGTATACACACACACCCCTATCGCTTCTCTTCAA 1557
QY 120 snThrGlnHisTyrHisTyrPleuArgProAsnSerGluLeuSerGlnIleArgPheThrI 140
Db 1558 ATACTCAGATTACCATTTGCTTGAAGGCGCAAAATTCAGAGCTTCTCAAAATCAGATTACA 1617
QY 140 leSerIlePheIleAsnGlyGluThrSerProSerHis 152
Db 1618 TCTCATTTTCATTACGGGGAAACATCCCCAGGCCAC 1655

RESULT 9
BC006319 2589 bp mRNA linear PRI 29-JUN-2004
LOCUS BC006319
DEFINITION Homo sapiens heat shock 27kDa protein family, member 7
(cardiovascular), mRNA (cDNA clone MGC:12642 IMAGE:4131506),
complete cds.
ACCESSION BC006319
VERSION BC006319.2 GI:33873500
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2589)
Straussberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Bueckow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

```

```

TITLE JOURNAL
PUBMED 12477932
REFERENCE 2 (bases 1 to 2589)
AUTHORS Straussberg, R.
TITLE Direct Submission
COMMENT Submitted (09-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NHI-MGC Project URL: http://mgc.nci.nih.gov
On Aug 19, 2003 this sequence version replaced gi:13623438.
Contact: MGC help desk
Email: cgaps-remail.nhi.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Ahter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kiong, P., Lurie, P., Legaspi, R.,
Maduro, Q.L., Mastello, C., Maskell, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantipod, S., Thomas, P.J., Touchman, J.W.,
Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/BLN at: http://image.llnl.gov
Series: IRAL Plate: 17 Row: 1 Column: 17
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 20070250.
Location/Qualifiers
  1..2589
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="MGC:12642 IMAGE:4131506"
    /tissue="Muscle, rhabdomyosarcoma"
    /clone_lib="NHI MGC 17"
    /lab_host="DH10B-R"
    /note="Vector: pOTB7"
  1..2589
    /gene="HSPB7"
    /note="synonym: CVASP"
    /db_xref="LocusID:27129"
    494..1006
    /gene="HSPB7"
    /codon_start=1
    /product="heat shock 27kDa protein family, member 7
(cardiovascular)"
    /protein_id="AA06319.1"
    /db_xref="GI:13623439"
    /db_xref="LocusID:27129"

```

/translation="MSHRTSSTRRAERFHSSSSSSSSSSSSSASRALPAODPPMEKA
LSMSDPSGFMFHPSEPLAPFAPPGAGNIKTLGDAYEFAVDVDRSPEDITVTISN
NHTEVRAEKLAADGTVMNTFAHKCOLPEDVDPTSVTSALREDSGLTIRARHPHTEHV
QOTFREIKI"

ORIGIN

Alignment Scores:

Pred. No.:	3,44e-50	Length:	2589
Score:	811.00	Matches:	149
Percent Similarity:	97.40%	Conservative:	1
Best Local Similarity:	96.75%	Mismatches:	2
Query Match:	95.64%	Indels:	2
DB:	9	Gaps:	0

US-09-977-418-8 (1-152) x BC006319 (1-2589)

QY 1 MetProHisLeuTyrTlleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
Db 1621 ATGCCACATCTGTATATAGATGGGGTTTTCACATACAGCTGTTCTGATTAACCTGCAT 1680
QY 21 GlnThrProAlaValIleuArgLeuLeuGlyProProGlyLysAlaLysTyrGlyTyrGly 40
Db 1681 GAAACTCTCGCGCTCTGCGCTCTGCGCTCTGCGCTCTGCGCTCTGCGCTCTGCGCTCTGCGG 1740
QY 41 TrrpGlyTrrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaPrometGlnThr 60
Db 1741 TGGGGTGTGCTCA 1800
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyr 80
Db 1801 GGATCACCCTAACAGAGATGGAGCCAGGCAATGGATGGGCTTTGGGCTCTCTCAAGCTGG 1860
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerle 100
Db 1861 ACCCAGCTCTTTCACCTTCCCTCCGCGAGTCAAGCTCTCATTCATCCCTCTTT 1920
QY 100 uUleTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProserA 120
Db 1921 AATCTAATGATCTATAGCTCGGTGTGTAAACACACACACCCCTATCGTTGCTCTCA 1980
QY 120 snThrGlnHisTyrHisTrrpLeuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
Db 1981 ATACTCAGCATTCACATTTGTTGAGCCCAATTCAGAGCTTTCTCAATTCAGATTTCAA 2040
QY 140 leSerIlePheIleAsnGlyGlnThrSerProSerHis 152
Db 2041 TCTCCATTTCATTAAACGGGGAACATCCCGAGCCAC 2078

RESULT 10

CQ783843

LOCUS CQ783843 2701 bp DNA linear PAT 17-MAR-2004

DEFINITION Sequence 3983 from Patent EP1396543.

VERSION CQ783843

KEYWORDS CQ783843.1 GI:45503747

SOURCE

ORGANISM Homo sapiens (human)

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

SOURCE

CDS

/note="unnamed protein product"

ORIGIN

Alignment Scores:

Pred. No.:	3.6e-50	Length:	2701
Score:	811.00	Matches:	149
Percent Similarity:	97.40%	Conservative:	1
Best Local Similarity:	96.75%	Mismatches:	2
Query Match:	95.64%	Indels:	2
DB:	6	Gaps:	0

US-09-977-418-8 (1-152) x CQ783843 (1-2701)

QY 1 MetProHisLeuTyrTlleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
Db 1766 ATGCCACATCTGTATATAGATGGGGTTTTCACATACAGCTGTTCTGATTAACCTGCAT 1825
QY 21 GlnThrProAlaValIleuArgLeuLeuGlyProProGlyLysAlaLysTyrGlyTyrGly 40
Db 1826 GAAACTCTCGCGCTCTGCGCTCTGCGCTCTGCGCTCTGCGCTCTGCGCTCTGCGG 1885
QY 41 TrrpGlyTrrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaPrometGlnThr 60
Db 1886 TGGGGTGTGCTCA 1945
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyr 80
Db 1946 GGATCACCCTAACAGAGATGGAGCCAGGCAATGGATGGGCTTTGGGCTCTCTCAAGCTGG 2005
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerle 100
Db 2006 ACCCAGCTCTTTCACCTTCCCTCCGCGAGTCAAGCTCTCATTCATCCCTCTTT 2065
QY 100 uUleTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProserA 120
Db 2066 AATCTAATGATCTATAGCTCGGTGTGTAAACACACACACCCCTATCGTTGCTCTCA 2125
QY 120 snThrGlnHisTyrHisTrrpLeuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
Db 2126 ATACTCAGCATTCACATTTGTTGAGCCCAATTCAGAGCTTTCTCAATTCAGATTTCAA 2185
QY 140 leSerIlePheIleAsnGlyGlnThrSerProSerHis 152
Db 2186 TCTCCATTTCATTAAACGGGGAACATCCCGAGCCAC 2223

RESULT 11

BD127861

LOCUS BD127861 2701 bp DNA linear PAT 18-SEP-2002

DEFINITION Primer for synthesizing full-length cDNA and use thereof.

VERSION BD127861

KEYWORDS BD127861.1 GI:23222806

SOURCE JP 2002017375-A/3292.

ORGANISM Homo sapiens (human)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

PI

/protein_id="CAF86886.1"
/db_xref="GI:45503748"
/translation="MQGLIHASLTAAHPTAPPGAGNIKTLGDAYEFAVDVDRSPED
IIVTSSNHIEVRAEKLAADGTVMNTFAHKCOLPEDVDPTSVTSALREDSGLTIRAR
HPHTEHVQOTFREIKI"

COMMENT

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 9, 2000 this sequence version replaced gi:6002162.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Ch1
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
RP3-336M4 is from the library RPci-3 constructed by the group of
Pieret de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC3

IMPORTANT: This sequence is not the entire insert of clone
RP3-336M4. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true right end of clone RP3-336M4 is at 76241 in this sequence.
The true right end of clone RP1-134019 is at 100 in this sequence.

FEATURES

```
source
1..76241
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="p36.11-36.2"
/clone="RP3-336M4"
/clone.lib="RPci-3"

repeat_region
2..174
/note="AluYb8 repeat: matches 1..174 of consensus"
complement(183..559)

misc_feature
/note="match: STS: Em:G37055 Em:NT5818"
609..935
/note="match: STS: Em:AL117303 Em:HS442J24S"
629..912
/note="match: GSS: Em:AQ264259"
634..852
/note="match: GSS: Em:B74942"
634..833
/note="match: GSS: Em:AQ311353"
634..768
/note="match: GSS: Em:AQ317636"
641..888
/note="match: GSS: Em:AQ678277"
complement(648..859)
/note="match: STS: Em:G17164"
648..813
/note="match: STS: Em:HS108ME1"
complement(648..785)
/note="match: GSS: Em:AQ867510"
650..927
/note="match: GSS: Em:AQ869729"

misc_feature
551..859
/note="match: GSS: Em:AQ754783"
652..877
/note="match: GSS: Em:AQ282607"
complement(654..814)
/note="match: GSS: Em:AQ118318"
658..935
/note="match: GSS: Em:AQ376808"
661..912
/note="match: STS: Em:G40890"
complement(664..931)
/note="match: GSS: Em:AQ455403"
complement(676..877)
/note="match: GSS: Em:AQ933387"
679..898
/note="match: STS: Em:G01739"
680..884
/note="match: GSS: Em:AQ589116"
complement(680..877)
/note="match: GSS: Em:AQ23541"
684..889
/note="match: GSS: Em:AQ378981"
687..850
/note="match: GSS: Em:AQ538621"
688..893
/note="match: GSS: Em:AQ50571"
complement(688..819)
/note="match: GSS: Em:AQ337044"
690..908
/note="match: GSS: Em:AQ200968"
690..850
/note="match: GSS: Em:AQ341026"
695..833
/note="match: STS: Em:G33512 Em:HS1045WH9"
complement(708..934)
/note="match: GSS: Em:AQ77257"
complement(708..834)
/note="match: GSS: Em:AQ064770"
708..831
/note="match: GSS: Em:AQ387275"
complement(728..936)
/note="match: STS: Em:AQ268772"
complement(745..911)
/note="match: GSS: Em:B64703"
complement(746..864)
/note="match: GSS: Em:AQ148978"
776..938
/note="match: GSS: Em:AQ984702"
789..922
/note="match: GSS: Em:AQ198586"
789..911
/note="match: GSS: Em:AQ30058"
1025..1133
/note="match: STS: Em:AQ30058"
1496..1581
/note="MIR repeat: matches 100..226 of consensus"
1850..2106
/note="L2 repeat: matches 2613..2709 of consensus"
complement(join(2349..2435,5293..5409))
/gene="dJ336M4.1"
/gene="dJ336M4.1"
complement(join(2349..2435,5293..5409))
/product="dJ336M4.1 (Miz-1)"
/note="match: CDNA: Em:U14556 Em:U23396 Em:Y09723"
match: ESTs: Em:AAQ21389 Em:U31841 Em:A155871 Em:AA624099
Em:AA78691 Em:AAQ18620 Em:AA408296 Em:AA294814
Em:AA432021 Em:AA481026"
/evidence="not_experimental"
2539..2582
/note="L2 repeat: matches 2648..2692 of consensus"
2583..2682
/note="L2 repeat: matches 2576..2680 of consensus"
2842..3213
repeat_region
```

```
/note="L2 repeat: matches 1731. .2159 of consensus"
repeat_region 3218. .3530
/note="AluJo repeat: matches 1. .312 of consensus"
repeat_region 3568. .3783
/note="AluJo repeat: matches 1. .196 of consensus"
repeat_region 3793. .3874
/note="L2 repeat: matches 2181. .2268 of consensus"
repeat_region 3883. .4185
/note="AluSx repeat: matches 1. .302 of consensus"
repeat_region 4490. .4709
/note="MIR repeat: matches 17. .260 of consensus"
misc_feature 4718. .5487
/note="CpG island"
evidence=not_experimental
repeat_region 5044. .5255
/note="106 copies 2 mer cc 58% conserved"
repeat_region 5576. .5794
/note="MIR repeat: matches 31. .262 of consensus"
repeat_region 5925. .6236
/note="L2 repeat: matches 373. .693 of consensus"
repeat_region 6321. .6632
/note="AluY repeat: matches 1. .309 of consensus"
repeat_region 6646. .6771
/note="PIRA_A repeat: matches 1. .133 of consensus"
repeat_region 6808. .6930
/note="L1PB2 repeat: matches 6032. .6155 of consensus"
repeat_region 6931. .7240
/note="AluY repeat: matches 1. .310 of consensus"
repeat_region 7241. .7930
/note="L1PB2 repeat: matches 5337. .6032 of consensus"
repeat_region 7931. .8240
/note="AluSx repeat: matches 1. .301 of consensus"
repeat_region 8241. .8966
/note="L1PB2 repeat: matches 4628. .5337 of consensus"
repeat_region 9024. .9084

Alignment Scores:
Pred. No.: 2.07e-48 Length: 76241
Score: 807.00 Matches: 148
Percent Similarity: 97.40% Conservative: 2
Best Local Similarity: 96.10% Mismatches: 2
Query Match: 95.17% Indels: 2
Gaps: 0

US-09-977-418-8 (1-152) x HSDJ36M4 (1-76241)
QY 1 MetProHisLeuTyriLeaSpGlyValPheProIleGlnLeuValArgGlyLeuHis 20
DB 44299 ATGCCACATCTGATATAGATGGGGTTTTCATACAGCTGGTTCGTATAACATGCAAT 44240
QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyValAlaLysTrpGlyTyrPgly 40
DB 44239 GAAACTCTCGCGCTCGCGCTGCTGGGCGCTCCAGGCAAGGCGACGTGGGGTGGGG 44180
QY 41 TrpGlyTyrPheSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 44179 TGGGGCTGCTCTCTCTCCCTCCAGGCGCTGTTTGGGGTGCTCCAGCAGACA 44120
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyr 80
DB 44119 GGATCACCATTACAGAGATGAGAGCAGGATGATGGGCTTTGGGCTCTCAAGTTGG 44060
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 44059 ACCCCAGCTTCTTCCACCTTCCCTCCGGGAGCTCAGCTCTCATTCATCCCTCTTT 44000
QY 100 uileTyrgIuSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSera 120
DB 43999 AATCTATGAAATCTATAGGCTCGTGTGTATACACACACACCCCTATCGTGTCTTCAA 43940
QY 120 snTrGlnHisTyrHisTrpLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrI 140
DB 43939 ATACTCAGCATTTACCATTTGGTTGGAGCCAAATTCAGAGCTTTCTCAATCAGATTTTCAA 43880
```

```
QY 140 leSerIlePheIleAsnGlyGlnThrSerProSerHis 152
DB 43879 TCTCATTTTCATTACCGGGAACATCCCGAGCGCAC 43842

RESULT 15
AX747447 2410 bp mRNA linear PAT 20-JUN-2003
LOCUS AX747447
DEFINITION Sequence 972 from Patent EP1308459.
ACCESSION AX747447
VERSION AX747447.1 GI:32131835
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 Isogai,T., Sugiyama,T., Otsuki,T., Makamatsu,A., Sato,H., Ishii,S.,
Yamamoto,O.,I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagaiari,K. and
Masuko,Y.
Full-length cDNA sequences
Patent: EP 1308459-A 972 07-MAY-2003;
Helix Research Institute (JP) ; Research Association for
Biotechnology (JP)

FEATURES
source location/Qualifiers
1..2410
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 1.46e-49 Length: 2410
Score: 802.00 Matches: 148
Percent Similarity: 96.75% Conservative: 1
Best Local Similarity: 96.10% Mismatches: 3
Query Match: 94.58% Indels: 2
Gaps: 0

US-09-977-418-8 (1-152) x AX747447 (1-2410)
QY 1 MetProHisLeuTyriLeaSpGlyValPheProIleGlnLeuValArgGlyLeuHis 20
DB 1478 ATGCCACATCTGATATAGATGGGGTTTTCATACAGCTGGTTCGTATAACATGCAAT 1537
QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyValAlaLysTrpGlyTyrPgly 40
DB 1538 GAAACTCTCGCGCTCGCGCTGCTGGGCGCTCCAGGCAAGGCGACGTGGGGTGGGG 1597
QY 41 TrpGlyTyrPheSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 1598 TGGGGCTGCTCTCTCTCCCTCCAGGCGCTGTTTGGGGTGCTCCAGCAGACA 1657
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyr 80
DB 1658 GGATCACCATTACAGAGATGAGAGCAGGCGATGATGGGCTTTGGGCTCTCAAGTTGG 1717
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 1718 ACCCCAGCTTCTTCCACCTTCCCTCCGGGAGCTCAGCTCTCATTCATCCCTCTTT 1777
QY 100 uileTyrgIuSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSera 120
DB 1778 AATCTATGAAATCTATAGGCTCGTGTGTATACACACACACCCCTATCGTGTCTTCAA 1837
QY 120 snTrGlnHisTyrHisTrpLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrI 140
DB 1838 ATACTCAGCATTTACCATTTGGTTGGAGCCAAATTCAGAGCTTTCTCAATCAGATTTTCAA 1897
QY 140 leSerIlePheIleAsnGlyGlnThrSerProSerHis 152
DB 1898 TCTCATTTTCATTACCGGGAACATCCCGAGCGCAC 1995
```

Search completed: October 24, 2004, 17:12:47
job time : 3591 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - nucleic search, using frame_plus_p2n model

Run on: October 24, 2004, 16:04:28 ; Search time 84 Seconds

(without alignments)
1286.189 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848

Sequence: 1 MRLPLVDYGVPIQLVREKLH.....LSQIRPTISIFINGERSPSH 152

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 824507 segs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seg length: 0

Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1h
-Q=/cgn2_1/USFTO.spool/US09977418/runat_22102004_115948_1048/app_query.fasta_1.327
-DB=Issued_Patents_NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOFCU=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR.SCORB=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTEXT=plo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USR=US09977418 @CGN 1.1 69 @runat_22102004_115948_1048 -NCPU=6 -ICPU=3
-NO MMAP -LARGEOBJSRV -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGIOG
-DEV.TIMEOUT=120 -MAIN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

1: /cgn2_6/ptodata/1/ina/5A.COMB.seg.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seg.*
3: /cgn2_6/ptodata/1/ina/5A.COMB.seg.*
4: /cgn2_6/ptodata/1/ina/5B.COMB.seg.*
5: /cgn2_6/ptodata/1/ina/PTCUS.COMB.seg.*
6: /cgn2_6/ptodata/1/ina/backfile1.seg.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	11.4	1330	US-10-329-668-17	Sequence 17, Appl
2	97	11.4	2000	US-08-016-434-1291	Sequence 1291, Ap
3	97	11.4	2000	US-09-919-497-10	Sequence 10, Appl
4	95.5	11.3	4403765	US-09-103-840A-1	Sequence 2, Appl
5	95.5	11.3	4411529	US-09-103-840A-1	Sequence 1, Appl
6	93.5	11.0	17000	US-09-691-220-3	Sequence 3, Appl
7	93	11.0	15202	US-09-602-787A-341	Sequence 341, Appl
8	93	11.0	15202	US-08-922-635-21	Sequence 21, Appl
9	92	10.8	13857	US-08-620-312D-75	Sequence 75, Appl
10	92	10.8	18609	US-08-943-731-1	Sequence 9, Appl
11	91.5	10.8	3992	US-09-944-807-9	Sequence 9, Appl
12	91	10.7	2733	US-09-799-451-480	Sequence 480, App

C 13	91	10.7	4480	3	US-09-191-171-7	Sequence 7, Appl
C 14	91	10.7	4480	3	US-09-385-707-7	Sequence 7, Appl
C 15	91	10.7	6238	4	US-09-638-686C-6	Sequence 6, Appl
C 16	91	10.7	8460	1	US-08-469-005A-9	Sequence 9, Appl
C 17	90	10.6	1512	3	US-08-955-918C-8	Sequence 8, Appl
C 18	90	10.6	1512	3	US-08-697-766A-8	Sequence 8, Appl
C 19	90	10.6	1785	3	US-08-264-578-7	Sequence 7, Appl
C 20	90	10.6	1801	3	US-08-955-918C-6	Sequence 6, Appl
C 21	90	10.6	1801	3	US-08-697-766A-6	Sequence 6, Appl
C 22	90	10.6	8519	3	US-09-012-072-1	Sequence 1, Appl
C 23	89.5	10.6	1218	3	US-09-120-601-1	Sequence 1, Appl
C 24	89.5	10.6	1218	3	US-08-081-610-1	Sequence 1, Appl
C 25	89.5	10.6	2205	1	US-09-016-454-1472	Sequence 1472, Ap
C 26	89	10.5	1719	4	US-08-948-569A-7	Sequence 7, Appl
C 27	89	10.5	1734	2	US-08-663-808-1	Sequence 1, Appl
C 28	89	10.5	1734	2	US-09-188-469-7	Sequence 7, Appl
C 29	89	10.5	1734	3	US-09-332-740-1	Sequence 1, Appl
C 30	89	10.5	1734	3	US-09-188-466-1	Sequence 7, Appl
C 31	89	10.5	1734	3	US-09-397-238A-7	Sequence 7, Appl
C 32	89	10.5	1734	3	US-09-368-282-1	Sequence 1, Appl
C 33	89	10.5	1734	4	US-09-566-708A-1	Sequence 1, Appl
C 34	89	10.5	2326	2	US-08-231-193A-41	Sequence 41, Appl
C 35	89	10.5	2326	2	US-08-486-473A-41	Sequence 41, Appl
C 36	89	10.5	2326	3	US-08-940-065A-41	Sequence 41, Appl
C 37	89	10.5	2326	3	US-08-940-035A-41	Sequence 41, Appl
C 38	89	10.5	2326	3	US-08-935-105A-41	Sequence 41, Appl
C 39	89	10.5	2326	4	US-09-648-797-41	Sequence 41, Appl
C 40	89	10.5	2326	4	US-09-386-123-41	Sequence 41, Appl
C 41	89	10.5	2326	4	US-08-231-193A-44	Sequence 44, Appl
C 42	89	10.5	3243	2	US-08-486-273A-44	Sequence 44, Appl
C 43	89	10.5	3243	2	US-08-486-273A-44	Sequence 44, Appl
C 44	89	10.5	3243	3	US-08-480-474-44	Sequence 44, Appl
C 45	89	10.5	3243	3	US-08-480-474-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-10-329-668-17
; Sequence 17, Application US/10329668
; Patent No. 6696473
; GENERAL INFORMATION:
; APPLICANT: Martin Richard
; APPLICANT: Brennon Todd
; APPLICANT: Kahl Jeffrey Dean
; APPLICANT: Wang Tie-Lin
; TITLE OF INVENTION: HETEROCYCLIC MODULATORS OF NUCLEAR RECEPTORS
; FILE REFERENCE: 38205-3001
; CURRENT APPLICATION NUMBER: US/10/329,668
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: 60/342,720
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1330
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (97)...(837)
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Genbank XM_042579
; DATABASE ENTRY DATE: 2002-02-06
; US-10-329-668-17

Alignment Scores:
Pred. No.: 2.38
Score: 97.00
Percent Similarity: 35.54%
Best Local Similarity: 26.51%
Query Match: 11.44%
Length: 1330
Matches: 44
Conservative: 15
Mismatch: 51
Indels: 56
Gaps: 10
DB:

US-09-977-418-8 (1-152) x US-10-329-668-17 (1-1330)

QY 2 ProHleuTyrlleaspolyValPheProIleGlnLeuValArgGluLysLeuHis--- 20
 Db 396 CCTTACTACTCTCTGCGGGA-----CAACAAGACTGCACAGT 434
 QY 21 ---GluThrProAla-----ValLeuArgLeuLeuGlyProProGlyLys--- 34
 Db 435 GGACAGCGCCGACCGGACCGCTGTACAGTACGCGCTTACAGAGTGCCTGCGCACAG 494
 QY 35 -----AlaLysTyrGlyTyr 39
 Db 495 CATGAAGAGGAGCGCGTACAGAGAGCGCTACCGCGGMAAGACAGAGATGGGAGATG 554
 QY 40 GlyTyrGlyTyrSerPheSerLeuProGlnAlaCys-----Val**GlyAlaIla 56
 Db 555 GGAAGGGGCTGGGGAGCCCGGAGAGATGCTGTGACAGAGATCTGGAGCGAGACT 614
 QY 57 PrometGlnThrGlySer---ProAsnArgAspGlySerGlnGlyMetAspGly----- 73
 Db 615 TGCTGTGAACAGAAAGATGACACAGGCGTGTGAGGGTCTGAGGGGACCGGGGAGCGG 674
 QY 74 -----AlaLeuGlyProArgGlyTyrThrProAla 83
 Db 675 CAGCAGCGTAGTGTGGGCTCAATCCACTCTCTGATGGGGGTGGGGAGAGCAG 734
 QY 84 SerCySHleuProleuArgGlnSerAlaLeuHisProSerProSerLeuIleTyrGlu 103
 Db 735 TGT-----AGTGTGTTCATATCCCTCCCTCTCTTCCCTCAT 776
 QY 104 SerIleGlySerValCysValThrThrHisProTyrArgCys-----ProSerAsnThr 121
 Db 777 AACCTTCCTAC-----ACTACTGTGAGATGAGAGTGTGCAAAACAGAGCTTTTC 827
 QY 122 GlnHisTyrHisTyrLeu 127
 Db 828 AACCATCTGAGTGCATG 845

RESULT 2

US-09-016-434-1291
 ; Sequence 1291, Application US/09016434
 ; Patent No. 6500938
 ; GENERAL INFORMATION:
 ; APPLICANT: Janice Au-Young
 ; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
 ; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
 ; NUMBER OF SEQUENCES: 1490
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 ; STREET: 3174 PORTER DRIVE
 ; CITY: PALO ALTO
 ; STATE: CALIFORNIA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/016,434
 ; FILING DATE: HEREWITH
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Zeller, Karen J.
 ; REGISTRATION NUMBER: 37,071
 ; REFERENCE/DOCKET NUMBER: PA-0002 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 855-0555
 TELEFAX: (650) 845-4166
 INFORMATION FOR SEQ ID NO: 1291:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2000 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: GENBANK
 CLONE: g29980
 US-09-016-434-1291

Alignment Scores:

Pred. No.: 4.2 Length: 2000
 Score: 97.00 Matches: 37
 Percent Similarity: 40.46% Conservative: 16
 Best Local Similarity: 28.24% Mismatches: 12
 Query Match: 11.44% Indels: 36
 DB: 4 Gaps: 10

US-09-977-418-8 (1-152) x US-09-016-434-1291 (1-2000)

QY 27 ArgLeuLeuGlyProProGlyLysAlaLysTyrGly-----TyrGlyTyr 41
 Db 184 AGACAAGAGACCGGAGCGCGCATCACGCTGGGGCGCTGAGCGCGGCATGTCATG 243
 QY 42 GlyTyrSer---PheSerLeuPro-----Gln 49
 Db 244 AAGTGGCAGCCCTGACGAGCGCTGAGCGCGCTGCGGGGAGCGAGCGCGCAT 303
 QY -50 AlaCysVal**GlyAlaIlaProMetGlnThrGlySerPro----- 63
 Db 304 GCGTGTGTGAGACTCGCGCTCTCTCTTTCATCCCGCCAGCATGCGCGCTG 363
 QY 64 AsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArg-----GlyTyr 80
 Db 364 TCAAGTGCCTTCAAGCACCATGCTGCGGCC--GGCCAGAGCGGCATGCGCTG 420
 QY 81 ThrProAlaSerCySHleuProleuArgGlnSerAlaLeuHisProSerProSerLeu 100
 Db 421 AGC--ACATCGTCCCAAGCGGAGCTCGGCGCGCTGCTGCGCGCGCTACACAG 477
 QY 101 IleTyrGlnSerIleGlySerValCysValThrThr-----HisProTyrArgCysPro 118
 Db 478 CCGTGG-----TGTTCGTGACGAGCGGACCGCGCGCTGACGCGCCA 522
 QY 119 SerAsnThrGlnHisTyrHisTyrLeuArgPro 129
 Db 523 AGCGCAGCGCACCCGCGCGCTG--CGGCGG 552

RESULT 3

US-09-919-497-10
 ; Sequence 10, Application US/09919497
 ; Patent No. 6773883
 ; GENERAL INFORMATION:
 ; APPLICANT: Mutter, George L.
 ; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
 ; FILE REFERENCE: B0801/725
 ; CURRENT APPLICATION NUMBER: US/09/919,497
 ; CURRENT FILING DATE: 2001-07-31
 ; PRIOR APPLICATION NUMBER: US 60/221,735
 ; PRIOR FILING DATE: 2000-07-31
 ; NUMBER OF SEQ ID NOS: 100
 ; SOFTWARE: Patentin version 3.0
 ; SEQ ID NO 10
 ; LENGTH: 2000
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-919-497-10
 Alignment Scores:

Pred. No.: 4.2 Length: 2000
Score: 97.00 Matches: 37
Percent Similarity: 40.46% Conservative: 16
Best Local Similarity: 28.24% Mismatches: 42
Query Match: 11.44% Indels: 36
DB: 4 Gaps: 10

US-09-977-418-8 (1-152) x US-09-919-497-10 (1-2000)

QY 27 ArgLeuLeuGlyProPogLYLysAlaLysTrpGly-----TrpGlyTrp 41
DB 184 AGACAAAGACGACCGCAGGCGCCATCAAGCTGGGGCGCTGAGCGCCGCGCATGCTATG 243
QY 42 GlyTrpSer--PheSerLeuPro-----Gln 49
DB 244 AAGTGGGACCTGTGACGCTGAGGCGCTGTGAGGAGAGAGCGGCGCAT 303
QY 50 AlaCysVal**GlyAlaAlaProMetGlnThrGlySerPro----- 63
DB 304 GCGCTGCTGCTGAGACTGCGCTCTTCTTCGCTTCAACGCGCGGCACATGCGGCTCTG 363
QY 64 AsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArg-----GlyTrp 80
DB 364 TCAACGCGCGCTTACAGACCATGTGCGGCGCC---GGGCCAAGGGCGCATGAGGCTG 420
QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
DB 421 AGC---ACATGTCGCGCCCAAGCGCGAGCTCGCGCGCTGCTGCGCGGCGCTTACACG 477
QY 101 IleTrpGluSerIleGlySerValCysValThr-----HisProTyrArgCysPro 118
DB 478 CCGTGG-----TGTGCTGTGACGACGAGCGCCGCGCTTGAAGCGGCGCA 522
QY 119 SerAsnThrGlnHisTyrHisTrpLeuArgPro 129
DB 523 AGCGCGAGCGCACCGCTGCGCTG---CGGCGG 552

RESULT 4
US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURES:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Alignment Scores:
Pred. No.: 2.4e+05 Length: 4403765
Score: 95.50 Matches: 36
Percent Similarity: 33.86% Conservative: 7
Best Local Similarity: 28.35% Mismatches: 37
Query Match: 11.26% Indels: 47
DB: 3 Gaps: 8

US-09-977-418-8 (1-152) x US-09-103-840A-2 (1-4403765)

QY 30 GlyProPogLYLysAlaLysTrpGlyTrpGlyTrpSerPheSerLeuProGln 49
DB 755982 GGGCGCTGCGCGCGCGGCGGCTGTGCTTCTTCTGTGTGTCTGTCTGCG-----CGC 755929
QY 50 AlaCysVal**GlyAlaAlaProMetGlnThrGlySerProAsnArgAspGlySerGln 69
DB 755928 CACTGTGCG-----AAGAGATCATCTACCGCG 755902
QY 70 GlyMetAspGlyAlaLeu-----GlyProArgGlyTrpThrPro 82
DB 755901 GCGTGTCTGCGGCTGTCTGTGATTCGCGCGCTGCGGCTGCGGCGCGCTGTGATGATACCA 755842
QY 83 Ala---SerCysHisLeuProLeuArgGlnSer---AlaLeuHis----- 95
DB 755841 CCGTGTCTTCCGATTCGCGCCCATCTCGAGTTGCGCCGAGCTTCACCTGTGNGTGTGTG 755782
QY 96 ProSerProSerLeuIleTyrGluSerIle----- 105
DB 755781 CCATCCGATCGCGCTGCGACGCTTCTATTCCGGGCTCTACTGCGCAGCATGTCACGC 755722
QY 106 -----GlySerValCys-----ValThrHisProTyr 115
DB 755721 ACCAAGTACCAACTATTGCGCGGATCGATTGCTGTGCTGTGCTGCTGCTGCGCGCATAT 755662
QY 116 ArgCysProSerAsnThrGln 122
DB 755661 CGTGTCCCTGAGCAGCGCGCA 755641

RESULT 5
US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37RV
US-09-103-840A-1

Alignment Scores:
Pred. No.: 2.41e+05 Length: 4411529
Score: 95.50 Matches: 36
Percent Similarity: 33.86% Conservative: 7
Best Local Similarity: 28.35% Mismatches: 37
Query Match: 11.26% Indels: 47
DB: 3 Gaps: 8

US-09-977-418-8 (1-152) x US-09-103-840A-1 (1-4411529)

QY 30 GlyProPogLYLysAlaLysTrpGlyTrpGlyTrpSerPheSerLeuProGln 49
DB 754022 GGGCGTCTGCGCGCGGCGGCTGTGCTTCTTCTGTGTGTCTGTCTGCG-----CGC 753969
QY 50 AlaCysVal**GlyAlaAlaProMetGlnThrGlySerProAsnArgAspGlySerGln 69
DB 753968 CACTGTGCG-----AAGAGATCATCTACCGCG 753942
QY 70 GlyMetAspGlyAlaLeu-----GlyProArgGlyTrpThrPro 82
DB 753941 GCGTGTCTGCGGCTGTCTGTGATTCGCGCGCTGCGGCTGCGGCGCGCTGTGATGATACCA 753882

PRIOR APPLICATION NUMBER: DE 19933006.9
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: DE 19940764.9
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19940765.7
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19940766.5
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19940830.0
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19940831.9
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19940832.7
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19940833.5
PRIOR FILING DATE: 1999-08-27
PRIOR APPLICATION NUMBER: DE 19941378.9
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: DE 19941379.7
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: DE 19941395.9
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: DE 19942077.7
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: DE 19942078.5
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: DE 19942079.3
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: DE 19942088.2
PRIOR FILING DATE: 1999-09-03
NUMBER OF SEQ ID NOS: 678
SEQ ID NO 341
LENGTH: 1986
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
FEATURE:
NAME/KEY: CDS
LOCATION: (101)..(1963)
OTHER INFORMATION: RXA00878
US-09-602-787A-341

Alignment Scores:
Pred. No.: 10.3 Length: 1986
Score: 93.00 Matches: 40
Percent Similarity: 32.93% Conservative: 14
Best Local Similarity: 24.39% Mismatches: 55
Query Match: 10.97% Indels: 56
Gaps: 5

US-09-977-418-8 (1-152) x US-09-602-787A-341 (1-1986)
QY 15 ValArgLysLeuHsGluThrProAlaValLeuArgLeuGlyProGlyLys 34
Db 796 TTAGGGAAGAAGTACGATTAACCGCGCATCCGTTGGTGGAGCGCATGATAC 855
QY 35 AlaLysTrpGlyTrpGlyTrpSerPheSerLeuProGlnAlaCysVal***Gly 54
Db 856 TCTCCGCGGCGC--GGCAGCTAATA-TCACTCTCCGCGCATTAACCTGATTTCT 911
QY 55 AlaLysProMetGlnThrGlySerProAsnArgAspGlySerGlnGlyMetAspGlyAla 74
Db 912 ACCGCACTTGACAGCGGCATTTACCATCGCGCATGATCA-----TGCCTATCCAGC 965
QY 75 LeuGlyPro----- 77
Db 966 TTGACACATGAGCGACAAACCGGTGATCATGATGACTACATGTCGACTCCGCGAG 1025
QY 78 -----ArgGlyTrpThrProAlaSerCysHisLeu 87
Db 1026 GCGCATGCGCGGCTCCCGGACACTTTCGAGGTCATGGCGCAGAGTGGAGCCACCG 1085
QY 88 ProLeuArgLysLeuAlaLeuHsProSerProSerLeuLeuGlyGlnSerIleGlySer 107

Db 1086 CCAATAGAGCTTGTGACGCCACCTCGCTCAGACATCCACGATCA----- 1136
QY 108 ValCysValThrHisProTyrArgCysProSerAsnThrGlnHisTrpLeu 127
Db 1137 -----GTCTGGGACAGCGCCGCGGCTCCCGCTGGAGAACACGAGATCG----- 1181
QY 128 ArgProAsnSerGlnLeuSerGlnIleArgPheThrIleSerIlePheIleAsnGlyGlu 147
Db 1182 -----AATTCAAA 1190
QY 148 ThrSerProSer 151
Db 1191 AGCTACCTTCG 1202
RESULT 8
US-08-922-635-21
Sequence 21, Application US/08922635A
Patent No. 6033871
GENERAL INFORMATION:
APPLICANT: PILLETZ, John E.
APPLICANT: IVANOV, Tina R.
TITLE OF INVENTION: DNA MOLECULES ENCODING IMIDALINE RECEPTIVE POLYPEPTIDES
FILE REFERENCE: AND POLYPEPTIDES ENCODED THEREBY
Patent No. 6033871
CURRENT APPLICATION NUMBER: US/08/922,635A
EARLIER FILING DATE: 1997-09-03
EARLIER APPLICATION NUMBER: 08/650,766
EARLIER FILING DATE: 1996-05-20
EARLIER APPLICATION NUMBER: 60/012,600
EARLIER FILING DATE: 1996-03-01
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 21
LENGTH: 15202
TYPE: DNA
ORGANISM: Homo sapiens
US-08-922-635-21

Alignment Scores:
Pred. No.: 174 Length: 15202
Score: 93.00 Matches: 38
Percent Similarity: 41.44% Conservative: 8
Best Local Similarity: 34.23% Mismatches: 32
Query Match: 10.97% Indels: 34
Gaps: 7

US-09-977-418-8 (1-152) x US-08-922-635-21 (1-15202)
QY 25 ValLeuArgLeuLeuGlyPro-----ProGlyLysAlaLysTrp---GlyTrpGlyTrp 41
Db 1532 ATCTTGCTCTCTAGTGAGGACACAGCCCTGCTGGCCCTGCTGGGCTGGGCTGG 1591
QY 42 GlyTrpSerPheSerLeuProGlnAlaCysVal***GlyAla-AlaProMetGlnThrG 61
Db 1592 GGG-----AGAGTGCGCAGACCTGCTTCACAG 1621
QY 61 YserProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGly----- 79
Db 1622 GGTACAGACACAGGAGGCGATGCTT-CTGAGGCT-----GGTCTCCGCGGGGAG 1674
QY 80 -----TTPThrProAlaSerCysHisLeuProLeuArgGlnSerAla 94
Db 1675 ACATGCGAGGGGTGCTGCTGCTGATGCGAGCTGTGCTTGCTGAGAGACTCCCAATT 1734
QY 94 u-----HisProSerProSerLeuIleTyrGlnSerIleGlySerValCysVal 110
Db 1735 GCTCTGATGCCACATTCAGCTCT-----CTAGAGACCGGAGGT 1776
QY 110 ThrThrHisProTyrArgCysProSerAsn 120
Db 1777 GTCTGACAGCGCCTGAGGCTCCCTGTGAC 1807

RESULT 9

US-09-620-312D-75/c
Sequence 75, Application US/09620312D
Patent No. 6569662
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yongshong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungqing
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: Polypeptides
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT FILING DATE: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: 09/488,725
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pc_fl_genes version 1.0
SEQ ID NO 75
LENGTH: 13857
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(13857)
US-09-620-312D-75

Alignment Scores:

Pred. No.:	192	Length:	13857
Score:	92.00	Matches:	47
Percent Similarity:	44.96%	Conservative:	11
Best Local Similarity:	36.43%	Mismatches:	37
Query Match:	10.85%	Indels:	35
DB:	4	Gaps:	9

US-09-977-418-8 (1-152) x US-09-620-312D-75 (1-13857)

QY 11 ProTLEGLhLeuValArgLulysLeuHISGIUTNr-ProAlaValLeuArgLeuLGI 30
DB 7481 CCTGGGGAGCGCTCCAGAGAGCGCTCACGGGAGAGCTGTCTCATCTTGGC 7422
QY 30 YP-ProGlyLyAlaLys-----TTPGly---TTPGlyT 41
DB 7421 TCCCCAGAGAGCGCTCCAGCGGCTCCGGGCTCCATCTGGTTCTGTGGGCTG 7362
QY 41 pGIYTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThG1 61
DB 7361 GGGG-----TCTCTTCT-----CTGCAGCTGTCACTCCGACGACAGACACA 7320
QY 61 YSerPro-----AsnArgAspGlySerGlnGlyMetAspGly---Al 74
DB 7319 GGGCCCATGGCGCTGGCCACCGTGAAGCGGTACAGCTTATCTGTATTATGCTGTGGC 7260
QY 74 aleuGlyProArgGlyTTPThProAlaSerCysHISLeuProLeuArgGlnSerAla 94
DB 7259 TCAGAGCCCGAGATGAGTG-AACCGCTTCCTGAGC-----CCTGCACACCTCATGCTT 7207
QY 94 whs---ProSerProSerLeuLeuLeuSerLeuSerLeuValCysValThThH1 113

DB 7206 GGGCTGCCCATCTCTCCATCTCTGTACTG-----ACAGGAA 7171
QY 113 sProTyArgCysProSerAsnThr 121
DB 7170 GTGCTCAACTGTCTCTCGGAACT 7146

RESULT 10

US-08-943-731-1
Sequence 1, Application US/08943731
Patent No. 6265157
GENERAL INFORMATION:
APPLICANT: PROCTOR, DARWIN J.
APPLICANT: SPOTILA, LORETTA D.
APPLICANT: DELTAS, CONSTANTINOS D.
APPLICANT: SEREDA, LARISA
APPLICANT: LARSON, ANDREA W.
APPLICANT: PACK, MICHAEL
APPLICANT: COLIGE, ALAIN
APPLICANT: EARLY, JAMES
APPLICANT: KORKKO, JARMO
APPLICANT: ALA-KORKKO, LEBNA, et al.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
NUMBER OF SEQUENCES: 666
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANTICH SCHWARZE JACOBS & NADEL, P.C.
STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
STREET: FLR.
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-7086

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,731
FILING DATE: 03-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: US 08/212,322
FILING DATE: 14-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/803,628
FILING DATE: 03-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: DOYLE LEARY Ph.D., KATHRYN
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9598-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1264
TELEFAX: 215-567-2991
TELEX: 831-494

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 18609 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-943-731-1

Alignment Scores:
Pred. No.: 290
Score: 92.00
Percent Similarity: 39.20%
Best Local Similarity: 32.80%
Query Match: 10.85%
DB: 3

Pred. No.:	290	Length:	18609
Score:	92.00	Matches:	41
Percent Similarity:	39.20%	Conservative:	8
Best Local Similarity:	32.80%	Mismatches:	31
Query Match:	10.85%	Indels:	45
DB:	3	Gaps:	10

QY 40 -----GlyTTPGlyTTP 43
DB 262 CCGGAGTGTAGGGGCGAGGCGAGGCGGAGTGTCTAGCGGCGAGGCGGCGGCTGG 203
QY 44 SerPheSer-LeuPro---GlnAlaCysVal**GlyAlaAlaPromeGlnTh-GlySe 62
DB 202 CAGTATAGTCTCTCCCTGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 146
QY 62 rProAsnArgAspGly-----SerGlnGlyMetAspG1 73
DB 145 CCGAGGCGAGTGTGTAGGAGTGTAGGAGTGTGTAGGAGTGTGTAGGAGTGTGTAGGAG 86
QY 73 yAlaLeuGly-----ProArgGlyTTPThrProAlaSerCysHisLeuProLeuArg 90
DB 85 GGACTTGGGCTGGGCTGGGCGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 26
QY 90 gGlnSerAlaLeuHisProSer 97
DB 25 TTGCTCC---TTGCTCTCTGCT 7
RESULT 13
US-09-191-171-7/c
Sequence 7, Application US/09191171
Patent No. 6149909 6143294
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
APPLICANT: ANSON, Donald S.
APPLICANT: ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENTS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
TITLE OF INVENTION: SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
City: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/191,171
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/494,104
FILING DATE: 23-JUN-1995
APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00611
FILING DATE: 12-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/084,254
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: DiGiulio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8978Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 4480 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-191-171-7
Alignment Scores:
Pred. No.: 50.3
Score: 91.00
Percent Similarity: 40.82%
Best Local Similarity: 32.65%
Query Match: 10.73%
DB: 3
Length: 4480
Matches: 32
Conservative: 8
Mismatch: 34
Indels: 25
Gaps: 5
US-09-977-418-8 (1-152) x US-09-191-171-7 (1-4480)
QY 30 GlyProGlyGlyAlaGlyTTPGly-----TTP----- 39
DB 1908 GGTCCGCTGTGTGTATATGGGGGTGTCCCGAATTGGGGAAGAGCTGGGATCT 1849
QY 40 -----GlyTTPGlyTTPSerPheSerLeuProGlnAla 50
DB 1848 GCTGGCGGAGCAGCTTCTCTCTCTCAGAGTGAAGAGTGGCGGAGCTT----- 1795
QY 51 CysVal**GlyAlaAlaPromeGlnThGlySerProAsnArgAspGlySerGlnGly 70
DB 1794 ---GGGGGGGGGGGGGGCTCAGACACACAGGAGTACGG---CGGCCGGGAGACAGCG- 1742
QY 71 MetAspGlyAlaLeuGlyProArgGlyTTPThrProAlaSerCysHisLeuProLeuArg 90
DB 1741 GCTGAGGGCGCA---GAACACCGGGGGGAGCGAGGCGGAGCGGAGCTTCTGTGGAG 1685
QY 91 GlnSerAlaLeuHisProSerProSerLeuLeuTyrGlnSerIleGlySerVal 108
DB 1684 CGAGATGTATCTCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1631
RESULT 14
US-09-385-707-7/c
Sequence 7, Application US/09385707
Patent No. 6238652
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
APPLICANT: ANSON, Donald S.
APPLICANT: ORSBORN, Annette M.
APPLICANT: NELSON, Paul V.
APPLICANT: CLEMENTS, Peter R.
APPLICANT: MORRIS, Charles P.
APPLICANT: HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
TITLE OF INVENTION: SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
STREET: 400 Garden City Plaza
City: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,707
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 06/494,104
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Digilio, Frank S.
 REGISTRATION NUMBER: 31,346
 REFERENCE/DOCKET NUMBER: 89782
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 516-742-4343
 TELEFAX: 516-742-4366
 TELEX: 230 901 SANS UR
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4480 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-09-385-707-7

Alignment Scores:
 Pred. No.: 50.3 Length: 4480
 Score: 91.00 Matches: 32
 Percent Similarity: 40.82% Conservative: 8
 Best Local Similarity: 32.65% Mismatches: 34
 Query Match: 10.73% Indels: 25
 DB: 3 Gaps: 5

US-09-977-418-8 (1-152) x US-09-385-707-7 (1-4480)

QY 30 GlyProProGlyLysAlaLysTyrGly-----Tyr----- 39
 DB 1908 GGTCCGCTCGTCTGTAAATGGGGTGTCCGGAACCTGGGGAAGCTCCGATCT 1849
 QY 40 -----GlyTyrGlyTyrPheSerPheSerLeuProGlnAla 50
 DB 1848 GCTGCGGACGACCTTCTCTCCGATGAGATGAGATGAGTCCGCGCACCT----- 1795
 QY 51 CysVal**GlyAlaAlaPrometGlnThrGlySerProAsnArgAspGlySerGlnGly 70
 DB 1794 ---GCGGGCGGGCGGCTTCCAGCACCAAGGCTCAGCG---CGGCCCGGGGACAGCG- 1742
 QY 71 MetAspGlyAlaLeuGlyProArgGlyTyrThrProAlaSerCysHisLeuProLeuArg 90
 DB 1741 GCTGAGGGGCGCA---GAACACCGGGCGGACGAGGGGCGAGCGGCGACCTTCTCTGTGAG 1685
 QY 91 GlnSerAlaLeuHisProSerProSerLeuLeuIleTyrGlnSerIleGlySerVal 108
 DB 1684 GGAAGATGATGTCAGCGCGACGCCCGCTCCCAAGTGAAGAGTGTGATCCGTC 1631

RESULT 15

US-09-639-696C-6/c
 Sequence 6, Application US/09639696C
 Patent No. 6524835

GENERAL INFORMATION:
 APPLICANT: SCOTT, Hamish S.
 ANSON, Donald S.
 ORSBORN, Annette M.

NELSON, Paul V.
 CLEMENTS, Peter R.
 MORRIS, Charles P.
 HOPWOOD, John J.

TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
 SEQUENCES ENCODING SAME

NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON PEABODY LLP
 STREET: 990 Stewart Avenue
 CITY: Garden City

STATE: New York
 COUNTRY: USA
 ZIP: 11530

COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/639,696C
 FILING DATE: 16-Aug-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/191,171
 FILING DATE: 13-NOV-1998
 APPLICATION NUMBER: US 08/494,104
 FILING DATE: 23-JUN-1995
 APPLICATION NUMBER: AU PK9490/91
 FILING DATE: 14-NOV-1991
 APPLICATION NUMBER: AU PCT/AU92/00611
 FILING DATE: 12-NOV-1992
 APPLICATION NUMBER: US 08/084,254
 FILING DATE: 07-JUL-1993

ATTORNEY/AGENT INFORMATION:
 NAME: Pokalsky, Ann R.
 REGISTRATION NUMBER: 34,697
 REFERENCE/DOCKET NUMBER: 2249/304
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 516-832-7552
 TELEFAX: 516-832-7555
 INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
 LENGTH: 6238 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 US-09-639-696C-6

Alignment Scores:

Pred. No.: 79.7 Length: 6238
 Score: 91.00 Matches: 32
 Percent Similarity: 40.82% Conservative: 8
 Best Local Similarity: 32.65% Mismatches: 34
 Query Match: 10.73% Indels: 25
 DB: 4 Gaps: 5

US-09-977-418-8 (1-152) x US-09-639-696C-6 (1-6238)

QY 30 GlyProProGlyLysAlaLysTyrGly-----Tyr----- 39
 DB 3666 GGTCCGCTCGTCTGTAAATGGGGTGTCCGGAACCTGGGGAAGCTCCGATCT 3607
 QY 40 -----GlyTyrGlyTyrPheSerPheSerLeuProGlnAla 50
 DB 3606 GCTGCGGACGACCTTCTCTCCGATGAGATGAGATGAGTCCGCGCACCT----- 3553
 QY 51 CysVal**GlyAlaAlaPrometGlnThrGlySerProAsnArgAspGlySerGlnGly 70
 DB 3552 ---GCGGGCGGGCGGCTTCCAGCACCAAGGCTCAGCG---CGGCCCGGGGACAGCG- 3500
 QY 71 MetAspGlyAlaLeuGlyProArgGlyTyrThrProAlaSerCysHisLeuProLeuArg 90
 DB 3499 GCTGAGGGGCGCA---GAACACCGGGCGGACGAGGGGCGAGGGGCGACCTTCTGTGAG 3443
 QY 91 GlnSerAlaLeuHisProSerProSerLeuLeuIleTyrGlnSerIleGlySerVal 108
 DB 3442 GGAAGATGATGTCAGCGCGACGCCCGCTCCCAAGTGAAGAGTGTGATCCGTC 3389

Search completed: October 24, 2004, 18:24:49
 Job time: 1649 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 24, 2004, 16:01:52 ; Search time 427 Seconds

(without alignments)
1868.646 Million cell updates/sec

Title: US-09-977-418-8
Perfect score: 848
Sequence: 1 MPELYIDGVPIQLVREKLM.....LSQIRFTISIFINGERSPSH 152

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=x1h
-O=/cgm2_1/usfto.spool/US09977418/runat.22102004.115947.1008/app.query.fasta.1.327
-DBN_genseq.23Sep04 -QFMT=faaetap -SUFTX=eng -MINMATCH=0.1 -LOOFCU=0
-LOOPEXT=0 -UNITS=b1ts -START=1 -END=-1 -MATRIX=blolsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US09977418 @CGN 1.1 470 @runat.22102004.115947.1008 -NCPU=6 -ICPU=3
-NO MAMP -LARGEJURY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEF TIMEOUT=120 -MAIN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N_Genseq.23Sep04:*

1: genseq1980s:*\n2: genseq1990s:*\n3: genseq2000s:*\n4: genseq2001as:*\n5: genseq2001bs:*\n6: genseq2002as:*\n7: genseq2002bs:*\n8: genseq2003as:*\n9: genseq2003bs:*\n10: genseq2003cs:*\n11: genseq2003ds:*\n12: genseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	846	99.8	1987	4	AAF27852 Human NOV
2	846	99.8	1987	9	ACD40257 Human bre
3	846	99.8	1987	11	ADMS6372 Human CDN
4	846	99.8	1987	12	ADf66734 CDNA enco
5	846	99.8	1987	12	Ad119771 Human NOV
6	846	99.8	1987	12	AD060244 Human NOV

7	817	96.3	2142	4	AA160565
8	817	96.3	2142	10	ADCC31919
9	817	96.3	2152	12	ADQ17578
10	817	96.3	4056	5	ADM20171
11	811	95.6	2192	2	AAZ33583
12	811	95.6	2220	5	ADM19231
13	811	95.6	2701	4	AAK94832
14	811	95.6	2701	12	ADL31950
15	811	95.6	2783	10	ADCC29984
16	811	95.6	3490	8	ACC46692
17	811	95.6	3503	12	ADQ22267
18	802	94.6	2410	10	ADB62818
19	644	75.9	400	9	ACH16935
20	578	68.2	1550	4	AAFS4984
21	512	68.4	305	4	AAE27801
22	512	68.4	305	9	ACD40305
23	512	68.4	305	11	ADMS6439
24	512	60.4	305	12	ADf66802
25	512	60.4	305	12	AD060312
26	502	59.2	305	12	AD119839
27	113.5	13.4	2347	6	ABA93765
28	110.5	13.0	4427	5	AAE79292
29	107	12.6	175737	6	ABR83571
30	107	12.6	175737	10	ADL13596
31	107	12.6	175737	12	ADQ18934
32	103.5	12.2	21010	4	AAK89247
33	103.5	12.2	21010	4	AA105888
34	103.5	12.2	21010	4	AB198452
35	103.5	12.2	21010	8	ABE74628
36	103.5	12.2	21010	10	ABZ68148
37	103.5	12.2	21024	4	AAK89248
38	103.5	12.2	21024	4	AA105889
39	103.5	12.2	21024	4	AB198453
40	103.5	12.2	21024	8	ABE74629
41	103.5	12.2	21024	10	ABZ68149
42	103	12.1	1701	12	ACH87372
43	103	12.1	2888	11	ADM02718
44	103	12.1	2954	4	AA197913
45	103	12.1	2954	4	AA198066

ALIGNMENTS

RESULT 1	AAE27852	AAE27852 strand; CDNA; 1987 BP.
XX	AAE27852;	
AC	AAE27852;	
XX		
DT	30-MAR-2001 (first entry)	
XX		
DE	Human NOV4 CDNA.	
XX		
XX	Human; NOVX; antiinflammatory; cytoprotective; neuroprotective;	
KW	cerebroprotective; immunomodulator; vulnery; vasotrophic; gene therapy;	
KW	hyperplasia; tumour; testecrosis; psoriasis; Dupuytren's contracture;	
KW	diabetes; Rheumatoid arthritis; cerebral oedema; Alzheimer's disease; ss.	
OS	Homo sapiens.	
XX		
PN	WO200075321-A2.	
XX		
PD	14-DEC-2000.	
XX		
PF	01-JUN-2000; 2000WO-US015303.	
XX		
PR	03-JUN-1999; 99US-0137322P.	
PR	16-MAR-2000; 2000US-0189810P.	
PR	22-MAR-2000; 2000US-0191158P.	
PR	30-MAR-2000; 2000US-0193086P.	
PR	31-MAY-2000; 2000US-00137322.	
XX		
PA	(CURA-) CUPAGEN CORP.	

XX Shinkets RA, Fernandes E, Herrman J, Vernet C;
 XX WPI; 2001-102403/11.
 DR P-PSDB; AAB61132.
 XX
 PT New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
 PT diagnostic marker, protein therapeutic, antibody or small molecule drug
 PT target for treating immune, proliferative and metabolic diseases and
 PT wound healing.
 XX
 PS Claim 8; Page 26-27; 194pp; English.
 XX
 CC The present sequence encodes a new isolated polypeptide (NOVX). The NOVX
 CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
 CC treating or preventing NOVX-associated disorders. They are also useful
 CC for determining the presence of or a predisposition to a disease
 CC associated with altered levels of the NOVX polypeptide or nucleic acid.
 CC These NOVX-associated disorders include hyperplasias, tumours,
 CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
 CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
 CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
 CC are especially useful in gene therapy. Specifically, NOVX is useful as a
 CC diagnostic marker or prognostic marker, protein therapeutic and antibody
 CC target or small molecule drug target to treat disorders in the immune
 CC response pathway, thyroid and metabolic diseases, bone metabolic
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive
 CC disorders), proliferative diseases, or tissue regeneration and
 CC development (e.g. wound healing or treatment of burns)
 CC
 XX
 SQ Sequence 1987 BP; 448 A; 613 C; 532 G; 386 T; 0 U; 8 Other;
 Alignment Scores:
 Pred. No.: 1,92e-57 Length: 1987
 Score: 846.00 Matches: 152
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.76% Indels: 0
 DB: Gaps: 0
 US-09-977-418-8 (1-152) x AAF27852 (1-1987)
 QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnIleuHis 20
 DB 991 ATGCCACATCTGTATATAGATGGGTTTTCATATACAGCTGTTCTGTAATACTGCAT 1050
 QY 21 GlnThrProAlaValIleuArgLeuGlnGlyProProGlyValAlaIleuTyrGly 40
 DB 1051 GAACCTCGTCCGCTCTGCGCTGCTGGGCGCTCCAGCAAGCGAGGGGTTGGGG 1110
 QY 41 TTPGlyTTPSerPheSerIleuProGlnAlaCysVal**GlyValAlaIleuThr 60
 DB 1111 TGGGGCTGTCTCTCTCCCTCCACAGGCTGTGTCNTGGGGCTGCTCCATGCAACA 1170
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValAlaLeuGlyProArgGlyTyr 80
 DB 1171 GGATTCCTTACACGATGGAAGCCAGGCGCATGAGTGGGGCTTTGGGTCTCGAGTTGG 1230
 QY 81 ThrProAlaSerCysHisIleuProLeuArgGlnSerAlaLeuHisProSerProSerIleu 100
 DB 1231 ACCCCAGCTTCTTGCACCTTCCCTCCGAGTCAAGCTCTCCATCCATCCCTCTTTA 1290
 QY 101 IleTyrGlySerIleGlySerValCysValThrThrHisProTyrArgCysProSerAsn 120
 DB 1291 ATTATGAACTTATAGGCTCGGTGTGTGTAACAACACCCCTTATCGTTGCTTCAAT 1350
 QY 121 ThrGlnHisTyrHisTyrLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrIle 140
 DB 1351 ACTCACCATTTACATGGTTGAGGCGCAATTGAGAGCTTTCATCAATCAGATTTTCAATC 1410
 QY 141 SerIlePheIleAsnGlyGlnThrSerProSerHis 152
 DB 1411 TCCATTTTCATTACCGGGAAACATCCCGAGGCCAC 1446

RESULT 2
 ACD40257
 ID ACD40257 standard; DNA; 1987 BP.
 XX
 AC ACD40257;
 XX
 DT 03-SEP-2003 (first entry)
 XX
 DB Human breast tumour associated protein 47-like polypeptide NOV4 DNA.
 XX
 KW Tissue typing; cancer; breast cancer; colon cancer; lung cancer; sarcoma;
 KW pancreatic cancer; uterine cancer; organ transplantation disorder; ds;
 KW cardiovascular disease; melanoma; atherosclerosis; diabetes mellitus;
 KW ischaemic heart disease; haemorrhage; peripheral vascular disease; gene;
 KW thrombosis; hypertension; systemic lupus erythematosus; haematopoiesis;
 KW tissue regeneration; wound healing; hyperproliferative disorder;
 KW psoriasis; neural disorder; Parkinson's disease; Alzheimer's disease;
 KW Huntington's disease; multiple sclerosis; amyotrophic lateral sclerosis;
 KW ALS; peripheral neuropathy; nervous system tumour; neurotoxin; tremor;
 KW neuropathy; acute brain injury; peripheral nerve trauma; human; NOVX;
 KW gene therapy; epilepsy; breast tumour associated protein 47.
 XX
 OS Homo sapiens.
 XX
 FN US2003027158-A1.
 XX
 PD 06-FEB-2003.
 XX
 PF 15-OCT-2001; 2001US-00977418.
 XX
 FR 03-JUN-1999; 99US-0137322P.
 FR 16-MAR-2000; 2000US-0189810P.
 FR 22-MAR-2000; 2000US-0191158P.
 PR 30-MAR-2000; 2000US-0193086P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Fernandes E, Herrman J, Vernet C;
 XX WPI; 2003-492028/46.
 DR P-PSDB; ABO23236.
 XX
 PT New nucleic acid sequence encoding a human breast tumor-associated
 PT protein 47-like polypeptide, useful for treating cardiovascular
 PT disorders, neural disorders, diabetes mellitus and cancers.
 XX
 PS Claim 8; Page 16-17; 100pp; English.
 XX
 CC The invention relates to a new isolated NOV4 nucleic acid. The nucleic
 CC acid is useful for identifying a compound that binds the nucleic acid.
 CC The nucleic acid is useful in gene therapy, in screening assays, in
 CC detection assays e.g. chromosomal mapping, cell and tissue typing and
 CC forensic biology, predictive medicine e.g. diagnostic assays, prognostic
 CC assays, monitoring clinical trials, and pharmacogenomics and methods of
 CC treatment including therapeutic and prophylactic. The nucleic acid is
 CC also useful for expressing NOVX protein. The nucleic acid is also useful
 CC to provide polynucleotide reagents e.g. labelled probes that are useful
 CC in an in situ hybridisation technique, for identifying a specific tissue
 CC (for example brain tissue) and for use in forensic science. The nucleic
 CC acid is also useful for mapping genes on a chromosome and thus locating
 CC gene regions associated with genetic disease, identifying an individual
 CC from a minute biological sample and to aid in forensic identification of
 CC biological sample. The nucleic acid is also useful for treating cancer,
 CC especially cancers of the breast, colon, lung, pancreas or uterus, or a
 CC melanoma or sarcoma. The nucleic acid is also useful for treating
 CC disorders related to organ transplantation, cardiovascular diseases,
 CC atherosclerosis, ischaemic heart disease, haemorrhage and systemic lupus
 CC erythematosus. NOVX protein encoded by the nucleic acid is useful for
 CC regulating haematopoiesis, for regeneration of bone, cartilage, tendon

CC ligament and/or nerve tissue growth or regeneration and for wound
CC healing. The nucleic acid is also useful for treating infections,
CC hyperproliferative disorders e.g. psoriasis, and neural disorders
CC including Parkinson's disease, Alzheimer's disease, Huntington's disease,
CC multiple sclerosis, amyotrophic lateral sclerosis (ALS), peripheral
CC neuropathy, tumours of the nervous system, exposure to neurotoxins, acute
CC brain injury, peripheral nerve trauma or injury and other neuropathies,
CC epilepsy, and/or tremors. The present sequence represents DNA encoding a
CC human breast tumour associated protein 47-like polypeptide
XX
SQ Sequence 1987 BP; 448 A; 613 C; 532 G; 386 T; 0 U; 8 Other;

QY 61 GlySerProAsnArgLeuArgPheGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
 Db 1171 GGATCACTTAACAGACAGATGAGACCCAGGCGATGATGGAGCTTGGGTCTCCAGAGTTGG 1230
 QY 81 ThrProAlaSerCySHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
 Db 1231 ACCCCAGCTCTTGGCCACTTCCCTCCGCGACGTCAAGCTTCCATCCATCCCTCTTTA 1290
 QY 101 IleTyrGluSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAsn 120
 Db 1291 ATCTAGATCAATCTAAGAGCTCGGTGTGTATCAACAACACCCCATGCTGTCTTCAAA 1350
 QY 121 ThrGlnHisTyrHisTSTPLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrIle 140
 Db 1351 ACTCAGCATTAACATGTTGATGAGGCCAAATTCAGAGCTTCTCAATCAGATTTCACATC 1410
 QY 141 SerIlePheIleAsnGlyGluThrSerProSerHis 152
 Db 1411 TCCATTTTCATTAACGGGAAACATCCCGAGCCAC 1446
 RESULT 5
 AD119771
 ID AD119771 standard; DNA; 1987 BP.
 AC AD119771;
 XX
 DT 22-APR-2004 (first entry)
 DE Human NOV4 DNA.
 XX
 KM Secreted protein; NOVX; diagnosis; metabolic disorder; diabetes; obesity;
 infection; anorexia; cancer; cardiovascular disease; hypertension;
 KM atherosclerosis; neurodegenerative disorder; Alzheimer's disease;
 KM Parkinson's disease; epilepsy; immune disorder; osteoarthritis;
 KM haematopoietic disorder; inflammatory skin disorder; asthma;
 KM dyslipidemia; neurogenesis; cell differentiation; cell proliferation;
 haematopoiesis; wound healing; angiogenesis; chromosome mapping;
 KM tissue typing; preventive medicine; pharmacogenomics; gene therapy;
 KM anorectic; cardiac; virologic; antibacterial; fungicide; protozoacide;
 KM neurotropic; neuroprotective; dermatological; human; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 991..1449
 FT /*tag= b
 FT /product= "Human NOV protein"
 FT /transl_except= (pos:1171..1221, aa:Gly-Pro)
 FT /transl_except= (pos:1144..1152, aa:Val-Gly)
 FT sig_peptide 991..1152
 FT /*tag= a
 FT mat_peptide 1153..1446
 FT /*tag= c
 FT /product= "Human mature NOV protein"
 XX
 US2004002134-A1.
 XX
 PD 01-JAN-2004.
 XX
 PF 15-OCT-2001; 2001US-00977819.
 XX
 XX 03-JUN-1999; 99US-0137322P.
 PR 16-MAR-2000; 2000US-0189810P.
 PR 22-MAR-2000; 2000US-019158P.
 PR 30-MAR-2000; 2000US-019308P.
 PR 03-MAY-2000; 2000US-0201388P.
 PR 31-MAY-2000; 2000US-00584411.
 XX
 PA (CURA-) CURAGEN CORP.
 PI Shimkets RA, Fernandes ER, Herrman JL, Vernet CAM;
 XX

DR WPI, 2004-070737/07.
 DR P-PSDB; AD119772.
 XX
 PT New NOVX nucleic acids encoding human KIAA0768 protein-like and human
 PT protein PRO-228 polypeptides, useful for treating NOVX-associated
 PT disorders.
 XX
 XX Disclosure; SEQ ID NO 7, 95pp; English.
 XX
 CC The present invention is based in part on the discovery of novel secreted
 CC and membrane-bound polypeptides and their encoding polynucleotides. The
 CC nucleic acids and polypeptides are collectively referred as NOVX. The
 CC invention is useful for treating, preventing and diagnosing diseases such
 CC as metabolic disorders, diabetes, obesity, infectious diseases such as
 CC viral, bacterial, fungal, helminthic and protozoal infections, anorexia,
 CC cancer, cardiovascular diseases such as hypertension and atherosclerosis,
 CC neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,
 CC epilepsy, immune disorders such as osteoarthritis, haematopoietic
 CC disorders, inflammatory skin disorders, asthma and various dyslipidemias.
 CC The invention is also useful as targets for the identification of small
 CC molecules that modulate or inhibit e.g. neurogenesis, cell
 CC differentiation, cell proliferation, haematopoiesis, wound healing and
 CC angiogenesis, as hybridisation probes, in chromosome mapping, tissue
 CC typing, preventive medicine and pharmacogenomics. The invention is also
 CC useful in gene therapy. The present sequence is human NOV DNA.
 XX
 SQ Sequence 1987 BP; 448 A; 615 C; 530 G; 386 T; 0 U; 8 Other;
 Alignment Scores:
 Pred. No.: 1,92e-57 Length: 1987
 Score: 846.00 Matches: 152
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.76% Indels: 0
 DB: 12 Gaps: 0
 US-09-977-418-8 (1-152) x AD119771 (1-1987)
 QY 1 MetProHsLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlyLeuHis 20
 Db 991 ATGCCACATCTGTATATGATGGGGTTTTCATACAGCTGGTTCGAGAAACTGAT 1050
 QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyIleValIleTyrGly 40
 Db 1051 GAAACTCTCGCGCTCTCGCTGCTGGGGCTCCAGGCAAGGCCAAGTGGGTTGGGG 1110
 QY 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaIleProMetGluThr 60
 Db 1111 TGGGCTGTGCTCTTCTCCCTCCACAGGCTGTGTCTTGGGGCTGCTCCCATGACACA 1170
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
 Db 1171 GGATCACTTAACAGACAGATGAGACCCAGGCGATGATGGAGCTTGGGTCTCCAGAGTTGG 1230
 QY 81 ThrProAlaSerCySHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
 Db 1231 ACCCCAGCTCTTGGCCACTTCCCTCCGCGACGTCAAGCTTCCATCCATCCCTCTTTA 1290
 QY 101 IleTyrGluSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAsn 120
 Db 1291 ATCTAGATCAATCTAAGAGCTCGGTGTGTATCAACAACACCCCATGCTGTCTTCAAA 1350
 QY 121 ThrGlnHisTyrHisTSTPLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrIle 140
 Db 1351 ACTCAGCATTAACATGTTGATGAGGCCAAATTCAGAGCTTCTCAATCAGATTTCACATC 1410
 QY 141 SerIlePheIleAsnGlyGluThrSerProSerHis 152
 Db 1411 TCCATTTTCATTAACGGGAAACATCCCGAGCCAC 1446
 RESULT 6
 ADO60244
 ID ADO60244 standard; DNA; 1987 BP.

```

XX ADO60244;
XX
XX 15-JUL-2004 (first entry)
XX
XX Human NOV4 DNA.
XX
XX Human; NOVX protein; cancer; hyperproliferative disease; cirrhosis;
XX keloid; psoriasis; tissue hypertrophy; osteoarthritis;
XX atherosclerotic plaque formation; haemorrhage; ischaemic heart disease;
XX renal disease; thrombosis; diabetes mellitus; hypertension;
XX hypothyroidism; severe combined immunodeficiency; SCID; infection;
XX malaria; candidiasis; autoimmune disorder; connective tissue disease;
XX multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
XX autoimmune pulmonary inflammation; Guillain-Barre syndrome;
XX autoimmune thyroiditis; myasthenia gravis; graft-versus-host disease;
XX autoimmune inflammatory eye disease; asthma; burn; incision; ulcer;
XX autoimmune disease; Alzheimer's disease; Parkinson's disease;
XX Huntington's disease; amyotrophic lateral sclerosis; Shy-Drager syndrome;
XX haematopoiesis; wound healing; tissue repair; antitumor;
XX antiinflammatory; gene; ds.
XX
XX Homo sapiens.
XX
XX Key location/Qualifiers
XX CDS 991..1149
XX FT /tag= b
XX FT /product= "NOVX protein"
XX FT /transl_except= (pos:1147..1149, aa:Xaa)
XX FT /note= "Xaa corresponds to any amino acid"
XX FT sig_peptide 991..1065
XX FT /tag= a
XX FT mat_peptide 1066..1146
XX FT /tag= c
XX FT /product= "Mature NOVX protein"
XX
XX US2003134430-A1.
XX
XX 17-JUL-2003.
XX
XX 15-OCT-2001; 2001US-00977751.
XX
XX 03-JUN-1999; 99US-0137322P.
XX 16-MAR-2000; 2000US-0189810P.
XX 22-MAR-2000; 2000US-0191158P.
XX 30-MAR-2000; 2000US-0193086P.
XX 03-MAY-2000; 2000US-0201388P.
XX 31-MAY-2000; 2000US-00584411.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinketsu RA, Fernandes E, Herrman J, Vernet C;
XX MPI, 2004-068928/07.
XX F-PSDB; ADO60245.
XX
XX Novel substantially purified NOVX polypeptide for treating severe
XX combined immunodeficiency, candidiasis, cancer, asthma, multiple
XX sclerosis, systemic lupus erythematosus.
XX
XX Disclosure; SEQ ID NO 7; 155pp; English.
XX
XX The invention relates to human NOVX polypeptides and polynucleotides.
XX NOVX sequences are useful in the treatment of cancer, hyperproliferative
XX diseases, cirrhosis, keloid, psoriasis, tissue hypertrophy,
XX osteoarthritis, atherosclerotic plaque formation, haemorrhage, ischaemic
XX heart or renal disease, thrombosis, diabetes mellitus, hypertension,
XX hypothyroidism, asthma, burns, incisions, ulcers, periodontal disease,
XX Alzheimer's disease, Parkinson's disease, Huntington's disease,
XX amyotrophic lateral sclerosis, Shy-Drager syndrome, immune deficiencies
XX and disorders such as severe combined immunodeficiency (SCID), bacterial
XX infection, viral infection e.g. herpes viral infection, protozoan
XX infection e.g. malaria, fungal infection e.g. candidiasis, autoimmune

```

```

CC disorders such as connective tissue disease, multiple sclerosis, systemic
CC lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary
CC inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin
CC dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease
CC and autoimmune inflammatory eye disease. The invention is useful in
CC regeneration and tissue growth of bone, cartilage, tendon, ligament,
CC haematopoiesis regulation, wound healing and tissue repair. Sequences of
CC the invention also exhibit antitumor and antiinflammatory activities. The
CC present sequence is human NOVX DNA.
XX
XX Sequence 1987 BP; 448 A; 613 C; 532 G; 386 T; 0 U; 8 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1,92e-57 Length: 1987
XX Score: 846.00 Matches: 152
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 99.76% Indels: 0
XX DB: 12 Gaps: 0
XX
XX US-09-977-418-8 (1-152) x ADO60244 (1-1987)
XX
XX 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
XX Db ATGCCACATCTGTATATAGATGGCGCTTTTCCATACACCTGCTGTGAAAACTGCAT 1050
XX
XX 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTrpGlyTyrGly 40
XX Db 1051 GAAACCTCTGCGCTCTGCGCTGCGGCGCTCCGACGSCCAAGCCAACTGGAGGGG 1110
XX
XX 41 TrpGlyTyrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
XX Db 1111 TGGGGCTGTGCTCTTCCCTCCACAGGCGCTGTTCNTGGGCGTCTCCATGCAGACA 1170
XX
XX 61 GlySerProAsnArgAspGlySerGlnGlyMetLeuAspGlyValLeuGlyProArgGlyTyrP 80
XX Db 1171 GGATCACTPAACAGATGAGATGAGCCAGCGCATGAGAGGGGCTTGGCTCTGAGAGTTGG 1230
XX
XX 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
XX Db 1231 ACCCCAGCTTCTTGCCACCTTCCCTCCGCGACGTACGCTCTCATCCACCCCTCTTTA 1290
XX
XX 101 IleTyrGluSerIleGlySerValCysValIThrThrHisProTyrArgCysProSerAsn 120
XX Db 1291 ATCTATGAATCTATAGCTGCTGCTGTGTATACAAACACACCCCTATGCTGTCTTCAAT 1350
XX
XX 121 ThrGlnHisTyrHisIleuLeuArgProAsnSerGluLeuSerGlnLeuArgPheThrIle 140
XX Db 1351 ACTCAGCATTAACCATTTGTTGAGGCCAAATTCAGAGCTTCTCAATTCAGATTACAAATC 1410
XX
XX 141 SerIlePheIleAsnGlyGluThrSerProSerHis 152
XX Db 1411 TCCATTTTCATTAAACGGGAAACATCCCGAGGCCAC 1446
XX
XX RESULT 7
XX AA160565
XX ID AA160565 standard; cDNA; 2142 BP.
XX
XX AC AA160565;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 4554.
XX
XX Human; noctropic; immunosuppressant; cyostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemostatic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX
XX Homo sapiens.
XX

```


PN WO20015312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US034263.
 XX
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-0048725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00596042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA,
 PI Zhou F, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR P-PSDB; AAM41409.
 PT Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX
 PS Claim 1; SEQ ID NO 4554; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (A157798-A161369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral neuropathies, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilization of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukemias and
 CC C.N.S. disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 CC
 XX
 SQ Sequence 2142 BP; 451 A; 700 C; 568 G; 423 T; 0 U; 0 Other;
 XX
 Alignment Scores:
 Pred. No.: 4.08e-55 Length: 2142
 Score: 817.00 Matches: 150
 Percent Similarity: 98.05% Conservative: 1
 Best Local Similarity: 97.40% Mismatches: 1
 Query Match: 96.34% Indels: 2
 DB: 4 Gaps: 0
 US-09-977-418-8 (1-152) x AAI60565 (1-2142)
 QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGluValSerHis 20
 DB 1198 ATGCCACATCTGTATATAGATGGGGTTTTCATTAACGCTGTCGTGATAACTGCAT 1257
 QY 21 GluThrProAlaValIleuArgLeuLeuGlyProProGlyLysAlaIleValTyrGly 40
 DB 1258 GAAATCTCTGCGCTCCGCTGCTGCGGCTCCAGGCAAGGCAAGTGGGGTGGGG 1117
 QY 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
 DB 1318 TGGGGCTGTCTCTCTCCCTCCACAGGCTGTGTCTTGGGGCTGTCTCCCATGCAACA 1377
 QY 61 GlySerProAsnArgAspGlySerGlnIleMetAspGlyAlaLeuGlyProArgGlyTyr 80
 DB 1378 GGAATCCTTAACAGATGAGAGCCAGGCGCATGATGGGGCTTTGGGCTCTCCAGGTTGG 1437

QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 DB 1438 ACCCAGCTTCTTGCCACCTTCCCTCCGGGAGGACACTCTCATTCATCCCTCTTT 1497
 QY 100 uileTygIuSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSera 120
 DB 1498 AATCTATGATCTATAGGCTCGGTGTGTATACACACACCCCATATGCTGTCTTCAA 1557
 QY 120 snThGlnHisTyrHisTyrLeuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
 DB 1558 ATACTCAGCATTTACATTCGTTGAGGCCAAATTCAGACTTCTTCAATTCAGATTACA 1617
 QY 140 leSerIlePheIleAsnGlyIuThrSerProSerHis 152
 DB 1618 TCTCATTTTCACTAACGGGAAACATCCCGAGCAC 1655
 RESULT 8
 ADG31919
 ID ADG31919 standard; cDNA; 2142 BP.
 XX
 AC ADG31919;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human novel cDNA contig sequence, SEQ ID NO:2001.
 XX
 XX Human; diagnostic; drug screening; forensics; gene mapping;
 XX biodiversity; assessment; Parkinson's disease; Alzheimer's disease;
 XX neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
 XX ulcers; osteoporosis; autoimmune disease; cancer;
 XX molecular weight marker; food supplement; antiparkinsonian; nootropic;
 XX neuroprotective; antianemic; anticoagulant; thrombolytic; vulnerrary;
 XX anticancer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
 XX gene therapy; chromosome 1p36.11-36.2; ss.
 OS Homo sapiens.
 XX
 XX WO2003029271-A2.
 XX
 XX 10-APR-2003.
 XX
 XX 24-SEP-2002; 2002WO-US030474.
 XX
 XX 24-SEP-2001; 2001US-0324631P.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
 PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Wang G;
 PI Haley-Vicente D, Drmanac RT;
 XX
 DR WPI: 2003-371981/35.
 DR P-PSDB; ADG32686.
 XX
 XX New polynucleotide and polypeptide useful for diagnosing, preventing or
 XX treating conditions such as neurodegenerative diseases, anemias, platelet
 XX disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 XX cancer.
 PT
 PT Example 2; SEQ ID NO 2001; 1185pp; English.
 XX
 XX The invention relates to 971 novel human cDNA sequences (ADG29919-
 XX ADG30889) and the polypeptides they encode (ADG30890-ADG31860). The
 XX invention also relates to nucleic acid sequences over 99% identical with
 XX the novel human cDNAs. The invention additionally encompasses expression
 XX vectors and host cells comprising a nucleic acid of the invention; the
 XX recombinant production of a polypeptide of the invention; an antibody
 XX against a polypeptide of the invention; a method of detecting
 XX polynucleotides or polypeptides of the invention; and methods of
 XX identifying a compound which binds to a polypeptide of the invention. The
 XX invention further discloses methods of preventing, treating or
 XX ameliorating a medical condition; kits comprising polynucleotide probes

CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (AD31861-AD31862) and the polypeptides encoded by the contigs (AD32828
CC -AD33399). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a human contig
CC sequence used in an example of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pcc_sequences.

XX SQ Sequence 2142 BP; 451 A; 700 C; 569 G; 423 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4,08e-55	Length:	2142
Score:	817.00	Matches:	150
Percent Similarity:	98.058	Conservative:	1
Best Local Similarity:	97.408	Mismatches:	1
Query Match:	96.348	Indels:	2
DB:	10	Gaps:	0

US-09-977-418-8 (1-152) x AD31919 (1-2142)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnIysLeuHis 20
DB 1198 ATGCCACATCTGATATAGATGGGCTTTTCCAAACACCTGGTTCGATATAACTGCAT 1257
QY 21 GluThrProAlaValLeuArgLeuGlyProProGlyLysValAlaValArgGlyTyrGly 40
DB 1258 GAAACCTCTGCTCTGCGCTGCGGCGCTCCAGGCAAGGCGAAGTGGGCTTGGGG 1317
QY 41 TrrGlyTrrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 1318 TGGGGCTGGTCTTCCCTCCACAGGCTGTGTTCTTGGGGCTGCCATCCAGACAGA 1377
QY 61 GlySerProAsnArgSpGlySerGlnGlyMetAlaSpGlyValAlaLeuGlyProArgGlyTyr 80
DB 1378 GATACCTTAACAGAGATGAAAGCCAGGCGCATGAGTGGGCTTGGGCTTCCAGGTTGG 1437
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 1438 ACCCCAGCTTCTGGCACCCTCCCTCCGCGGCAAGTCAAGTCCATCCATCCCTCTTT 1497
QY 100 UleUtyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA 120
DB 1498 AATCATGATCATTAAGCTCGGTGTGTATACACACACACCCCATGTTGCTTTAA 1557
QY 120 snThcGlnHisTyrHisTrrPleuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
DB 1558 AATACAGCATTAACATGCTTGAAGGCCAAATTCAGAGCTTCTCAATTCAGATTACA 1617
QY 140 lAsertIlePheIleAsnGlyGlnThrSerProSerHis 152
DB 1618 TCTCCATTTTCATTAACGGGGGAAACATCCCGAGCCAC 1655
RESULT 9
AD317578
ID AD317578 standard; DNA; 2152 BP.
AC AD317578;
XX
XX 26-AUG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 395.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX ds.

XX Homo sapiens.

XX WO2004048938-A2.

XX 10-UDN-2004.

XX 26-NOV-2003; 2003WO-US038193.

XX 26-NOV-2002; 2002US-0429739P.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Aziz N, Ginsburg WM, Zlotnik A;

XX WPI; 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression
XX of a gene in a first soft tissue sample and a normal soft tissue sample
XX and comparing the gene expression, also useful in treating soft tissue
XX sarcoma.

XX Example 2; SEQ ID NO 395; 210bp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma
XX which comprises obtaining a first soft tissue sample from an individual
XX and a normal soft tissue sample from the same or different individual,
XX determining the expression of a gene in both samples and comparing the
XX expression of the gene in both soft tissue samples, where a higher level
XX of protein expression in the first soft tissue sample indicates the
XX presence of soft tissue sarcoma. The method of the invention has
XX cyostatic applications and may be useful for detecting soft tissue
XX sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX acid sequences may be useful in diagnostic and screening applications.
XX The current sequence is that of a human soft tissue sarcoma-upregulated
XX DNA of the invention. The current sequence is not shown within the
XX specification per se but was submitted in CD format by the inventor.

XX SQ Sequence 2152 BP; 459 A; 700 C; 569 G; 424 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.1e-55	Length:	2152
Score:	817.00	Matches:	150
Percent Similarity:	98.058	Conservative:	1
Best Local Similarity:	97.408	Mismatches:	1
Query Match:	96.348	Indels:	2
DB:	12	Gaps:	0

US-09-977-418-8 (1-152) x AD317578 (1-2152)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnIysLeuHis 20
DB 1201 ATGCCACATCTGATATAGATGGGCTTTTCCAAATCACTGTTCTGATATAACTGCAT 1260
QY 21 GluThrProAlaValLeuArgLeuGlyProProGlyLysValAlaValArgGlyTyrGly 40
DB 1261 GAAACCTCTGCTCTTCCCTCCACAGGCTGTGTTCTTGGGGCTGCCATCCATGACAGA 1380
QY 41 TrrGlyTrrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 1321 TGGGGCTGGTCTTCCCTCCACAGGCTGTGTTCTTGGGGCTGCCATGACAGA 1380
QY 61 GlySerProAsnArgSpGlySerGlnGlyMetAlaSpGlyValAlaLeuGlyProArgGlyTyr 80
DB 1381 GATACCTTAACAGAGATGAAAGCCAGGCGCATGAGTGGGCTTGGGCTTCCAGGTTGG 1440
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100

17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249219P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249248P.
PR 17-NOV-2000; 2000US-0249249P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251888P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251473P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251899P.
PR 08-DEC-2000; 2000US-0251907P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-476159/51.
XX
XX Isolated nucleic acid molecule encoding a channel/transporter protein is
PT used in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure; SEQ ID NO 978; 809pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule encoding a
CC channel/transporter protein or sequences at least 95% identical to a
CC these. The nucleic acids and proteins encoded by them are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. The antibodies to the proteins can also be used
CC in alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiodysplasia,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities. This sequence corresponds to an alternative version
CC of the gene of the invention.
XX
XX Sequence 4056 BP; 800 A; 1323 C; 1182 G; 751 T; 0 U; 0 Other;

Best Local Similarity: 97.40% Mismatches: 1
Query Match: 96.34% Indels: 2
DB: 5 Gaps: 0
us-09-977-418-8 (1-152) x ADM20171 (1-4056)
Qy 1 MetProHStLeuYrTlLeaSPGlyAlaPheProIleGLeuValaArgGlyLeuHis 20
Db 3075 ATGCCACATCTGTATATAGATGGGGTTTTCACATCAGCTGGTTCGTATTAACATGCAT 3134
Qy 21 GluThrProAlaValLeuAArgLeuLeuGlyProProGlyLyAlaLysTrpGlyTyrGly 40
Db 3135 GAACCTCCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 3194
Qy 41 TrpGlyTyrSerPheSerLeuProGlnAlaCysVal***GlyAlaLysProMetGlnThr 60
Db 3195 TGGGGCTGCTCTCTCTCCCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 3254
Qy 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
Db 3255 GGATCACCCTTAACAGATGGAAGCAGGCGATCGGGCTTGGGCTCTCGAGGTTGG 3314
Qy 81 ThrProAlaSerCysHisLeuPProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
Db 3315 ACCCCAGCTTCTGCGCACCCTTCCCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCTT 3374
Qy 100 UleTyrGluSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA 120
Db 3375 AATCTATGAATCTATAGCTCGGGTGTGTATACACACACCCCTATCGTTCCTTCA 3434
Qy 120 snThrGlnHisTyrHisTrpLeuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
Db 3435 ATACTCAGCATTAACATGTTGAGGCCAAATTCAGAGCTTCTCAATCAGATTTCACA 3494
Qy 140 IeSerIlePheIleAsnGlyGluThrSerProSerHis 152
Db 3495 TCTCCATTTCATTAAAGGGGAAACATCCCGAGCCAC 3532
RESULT 11
AAZ33583 ID AAZ33583 standard; cDNA; 2192 BP.
XX AAZ33583;
XX
XX 08-DEC-1999 (first entry)
XX
XX Human breast tumour-associated EST 43.
XX
XX Expressed sequence tag; EST; human; breast; cancer; cytostratic;
XX
XX medicaments; gene therapy; treatment; fat metabolism; ss.
XX
XX Homo sapiens.
XX
XX DE19813835-A1.
XX
XX 23-SEP-1999.
XX
XX 20-MAR-1998; 98DE-01013835.
XX
XX 20-MAR-1998; 98DE-01013835.
XX
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
XX Specht T, Hinemann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;
XX
XX WPI; 1999-528979/45.
XX
XX P-PSDB; AAY48501, AAY48502.
XX
XX Human nucleic acid sequences and protein products from normal breast
PT tissue, useful for breast cancer therapy.
XX
XX Claim 1a; 131; 206bp; German.

CC This invention describes novel human nucleic acid sequences from normal
CC breast tissue which have cytoskeletal activity. The nucleic acid sequences
CC can be used to produce and isolate full-length gene sequences. They can
CC be used to express proteins, which can be used as tools to find an
CC activity against breast cancer. The sequences can be used in sense or
CC antisense form. They are especially useful for medicaments for gene
CC therapy to treat breast cancer and for treating illnesses associated with
CC fat metabolism. AA33541-233610 represent expressed sequence tags
CC described in the method of the invention

XX Sequence 2192 BP; 457 A; 717 C; 588 G; 430 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1,25e-54	Length:	2192
Score:	811.00	Matches:	149
Percent Similarity:	97.40%	Conservative:	1
Best Local Similarity:	96.75%	Mismatches:	2
Query Match:	95.64%	Indels:	2
DB:	2	Gaps:	0

US-09-977-418-8 (1-152) x AA33583 (1-2192)

QY 1 MepProHisLeuTYrIleAspGlyValPheProIleGlnLeuValArgGlnLeuHis 20
DB 1198 ATGCCACATCTGTATAGATGGGCTTTTCCATACAGCTGTCTGTATACATGCAAT 1257
QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyIysAlaIysTTPGly 40
DB 1258 GAAACCTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGG 1217
QY 41 ThrGlyThrPheSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaPheMetGlnThr 60
DB 1318 TGGGGCTGTCCT 1377
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTTP 80
DB 1378 GGATCACCTACACAGATGAGAGGACGAGGATGATGGGCTTGGCTCTCGAGGTTGG 1437
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSer 100
DB 1438 ACCCAGCTCTTCTGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTT 1497
QY 100 uileTyGlnSerIleGlySerValCysValThr-ThrHisProTyArgCysProser 120
DB 1498 AATATAGATCTATAGCTCTGCTGTGTATACACACACACCCCTATCGTTGCTTCAA 1557
QY 120 snThrGlnHisTYrHisTTPLeuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
DB 1558 ATACTCGACATATACATGATGTTGAGGCCAATTAGAGCTTTCTCAATACAGATTACA 1617
QY 140 IeserIlePheIleAsnGlyGlnThrSerProserHis 152
DB 1618 TCTCCATTTTCAITTAACGGGAAACATCCCGAGCCAC 1655

RESULT 12

ID ADM19231 standard; cDNA; 2220 BP.

XX ADM19231;

XX 20-MAY-2004 (first entry)

XX Novel human channel/transporter gene #28.

XX ds; gene; immunosuppressive; antiarthritic; antineumatic;
XX antiproliferative; cytostatic; cardiatic; vasotropic; cerebroprotective;
XX neurotropic; neuroprotective; antibacterial; virucide; fungicide;
XX optalmological; gene therapy; channel/transporter protein;
XX rheumatoid arthritis; neoplasia; cardiac arrest; cerebrovascular disorder;
XX cerebral ischemia; angiogenesis; nervous system disorder;
XX Alzheimer's disease; ocular disorder; corneal infection; wound healing;
XX epithelial cell proliferation; skin aging; sunburn; transplantation;
XX chemotaxis; food additive.

XX Homo sapiens.
OS MO200154472-A2.
PN 02-AUG-2001.
PF 17-JAN-2001; 2001WO-US001307.
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214866P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 17-JUL-2000; 2000US-0216800P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225577P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228987P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232197P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.

PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0241826P.
 PR 08-NOV-2000; 2000US-0246174P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249318P.
 PR 17-NOV-2000; 2000US-0249344P.
 PR 17-NOV-2000; 2000US-0249345P.
 PR 17-NOV-2000; 2000US-0249346P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251303P.
 PR 05-DEC-2000; 2000US-0251388P.
 PR 05-DEC-2000; 2000US-0251479P.
 PR 05-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251890P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX

PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI, 2001-476159/51.
 DR P-PSDB; ADM19710.
 XX
 PT Isolated nucleic acid molecule encoding a channel/transporter protein is
 XX used in preventing, treating or ameliorating a medical condition.
 PS Claim 1; SEQ ID NO 38; 809gp; English.
 CC The invention relates to an isolated nucleic acid molecule encoding a
 CC channel/transporter protein or sequences at least 95% identical to a
 CC these. The nucleic acids and proteins encoded by them are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. The antibodies to the proteins can also be used
 CC in alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiodysplasia,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
 CC The polypeptides can also be used to aid wound healing and epithelial
 CC cell proliferation, to prevent skin aging due to sunburn, to maintain
 CC organs before transplantation, for supporting cell culture of primary
 CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
 CC also be used as a food additive or preservative to increase or decrease
 CC storage capabilities. This sequence corresponds to a gene of the
 CC invention.
 XX
 SQ Sequence 2220 BP; 461 A; 718 C; 597 G; 429 T; 0 U; 15 Other;
 Alignment Scores:
 Pred. No.: 1,27e-54 Length: 2220
 Score: 811.00 Matches: 149
 Percent Similarity: 97.40% Conservative: 1
 Best Local Similarity: 96.75% Mismatches: 2
 Query Match: 95.64% Indels: 0
 DB: 5 Gaps: 0
 US-09-977-418-8 (1-152) x ADM19231 (1-2220)
 QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
 DB 1252 ATGCCACATCTGTATATAGATGGGGCTTTCCAAATACAGCTGGTGTGATAACACGCAT 1311
 QY 21 GluThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTyrPgiLys 40
 DB 1312 GAATCTCTCCGCTCCGCGCCCTGCGGGCCCTCCAGGCAAGGCCAMGTGGGTTGGGG 1371
 QY 41 TTPGlyTyrPsePheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
 DB 1372 TGGGGCTGTCTTCTCCCTCCACAGGCTGTGTTCTTGGGGCTCTCCATGCGAGACA 1431
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
 DB 1432 GGATCACCCTAACAGATGGAGCCAGGGCATGAGATGGGGCTTTGGGCTCTCGAGGTTGG 1491
 QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 DB 1492 ACCCCAGCTTCTTGCCACACCTTCCCTCCGGGCGAGTCAGCTTCCATCCATCCCTCTTT 1551
 QY 100 uileTyrGlnSerIleGlySerValGysValThr-ThrHisProTyrArgCysProSerA 120
 DB 1552 AATCTATGAATCTATAGGCTGGGTGTGTGAACACACACCCCTATCGTGTCTTCA 1611
 QY 120 snThrGlnHisTyrHisTyrPleuArgProAsnSerGlnLeuSerGlnIleArgPheThrI 140

Db 1612 ATACTGACATTCATTCAGGAGGCAATTCAGAGCTTCTCAATCAGATTACAA 1671
Qy 140 |leSerIlePheIleAsnGlyGluThrSerProSerHis 152
Db 1672 TCTCCATTTTCATTACGGGGAACATCCCGAGCCAC 1709
RESULT 13
ID AAK94832 standard; cDNA; 2701 BP.
XX
AC AAK94832;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 3983.
XX
KM Human, full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX
PA (HELI-) HELIX RES INST.
PI Oca T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H,
XX
XX WPI; 2001-524255/58.
DR P-PSDB; AAM93873.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX
XX Claim 8; SEQ ID NO 3983; 1380bp + Sequence Listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesizing the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a full length human cDNA of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in CD-ROM format directly
CC from EPO
XX
SQ Sequence 2701 BP; 557 A; 867 C; 771 G; 506 T; 0 U; 0 Other;

Alignment Scores:
Pred. NO: 1.58e-54 length: 2701
Score: 811.00 Matches: 149
Percent Similarity: 97.40% Conservative: 1
Best Local Similarity: 96.75% Mismatches: 2
Query Match: 95.64% Indels: 2
DB: 4 Gaps: 0

US-09-977-418-8 (1-152) x AAK94832 (1-2701)

Qy 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
Db 1766 ATGCCACATCTGTATATAGATGGGCTTTTCATATACAGCTGGTTCGTATATAACGCAAT 1825
Qy 21 GlnThrProAlaValIleuArgLeuLeuGlyProProGlyLysAlaLysTyrGlyTyrGly 40

Db 1826 GAAATCTCTGCGCTCTCGCTGCTGGGGCTTCAGGCAAGGCGCACCTGGGGTTGGGGG 1885
Qy 41 TTPGlyTTPSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
Db 1886 TGGGGCTGCTCTTCTCTCCACAGGCTGTGTCTTGGGGGCTGCTCCATGCGAGCA 1945
Qy 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
Db 1946 GGATCACCCTAACAGAGATGAGAGCCAGGCGATGATGGGGCTTGGGCTCTCGAGGTTGG 2005
Qy 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
Db 2006 ACCCCAGCTTCTTGGCAGCTTCCCTCCGGGAGCAGCTCTCCATCCATCCCTCTTT 2065
Qy 100 vIleTyrGluSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSera 120
Db 2066 AATCTATGATATATAGGCTCGTGTGTGTATACACACACACCCCTATGCTGTCTTCA 2125
Qy 120 sTrThrGlnHisTyrHisTrpLeuArgProAsnSerGlyLeuSerGlnIleArgPheThrI 140
Db 2126 ATACTGACATTCATTCAGGAGGCAATTCAGAGCTTCTCAATCAGATTACAA 2185
Qy 140 |leSerIlePheIleAsnGlyGluThrSerProSerHis 152
Db 2186 TCTCCATTTTCATTACGGGGAACATCCCGAGCCAC 2223
RESULT 14
ID ADL31950 standard; cDNA; 2701 BP.
XX
AC ADL31950;
XX
DT 20-MAY-2004 (first entry)
XX
DE Full length human cDNA clone SeqID 3983.
XX
XX human; medicine; signal transduction; glycoprotein; transcription;
XX oligo-capping method; ss; gene.
XX
XX Homo sapiens.
XX
PN EP1396543-A2.
XX
PD 10-MAR-2004.
XX
PF 07-JUL-2000; 2003EP-00025638.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
PI Oca T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H,
XX
XX WPI; 2004-204755/20.
DR P-PSDB; ADL31951.
XX
PT New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT length human cDNAs.
XX
XX Example 1; SEQ ID NO 3983; 1340bp; English.
XX
CC This invention relates to a novel primers useful for synthesizing full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3'

CC ends using the oligo-capping method. This polynucleotide sequence is a
 CC full length human cDNA clone of the invention.

SO Sequence 2701 BP; 557 A; 867 C; 771 G; 506 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.58e-54	Length:	2701
Score:	811.00	Matches:	149
Percent Similarity:	97.40%	Conservative:	1
Best Local Similarity:	96.75%	Mismatches:	2
Query Match:	95.64%	Indels:	2
DB:	12	Gaps:	0

US-09-977-418-8 (1-152) x ADL1950 (1-2701)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
 Db 1766 ATCCACATCTGTATATAGATGGCGTTTTCATATACACTGGTGTGATTAACGCAAT 1825
 QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTyrPgly 40
 Db 1826 GAAATCTCCGCTCTCCCTGCGCTGCGGCTCCAGGCAAGCCACGCGGGGTTGGGG 1885
 QY 41 TrpGlyTyrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
 Db 1886 TGGGGCTGTCCTTCCCTCCACAGCGCTGTGTTCTTGGGGCTGCTCCATGACAGCA 1945
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
 Db 1946 GGATCACTCAACAGAAATGGAAGCCAGGCAATGATGAGGGGCTTGGGCTCTGAGAGTTGG 2005
 QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
 Db 2006 ACCCGAGCTTCTTCCACCTTCCCTCCGCGAGCTGCTCCATCCATCCCTCTTT 2065
 QY 100 ValTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSera 120
 Db 2066 AATCTATGATCATATAGCTCGCTGCTGTATACACACACACCCCTATGCTGCTTCAA 2125
 QY 120 snThrGlnHisTyrHisTSTPLeuArgProAsnSerGlnLeuSerGlnIleArgPheThr 140
 Db 2126 ATTCATGATCATACCATGTTGAGGCGCAAAATTCAGAGCTTCTCAATCATGATTTACA 2185
 QY 140 IseSerIlePheIleAsnGlyGlnThrSerProSerHis 152
 Db 2186 TCTCCATTTTCATTTACGGGGAACATCCCGAGGCAC 2223
 RESULT 15
 ADL29984
 ID ADL29984 standard; cDNA; 2783 BP.
 AC ADL29984;
 XX 18-DEC-2003 (first entry)
 DT
 XX Human novel cDNA sequence, SEQ ID NO:66.
 DE
 XX Human; diagnostic; drug screening; forensics; gene mapping;
 XX biotechnology assessment; Parkinson's disease; Alzheimer's disease;
 XX neurodegenerative diseases; anemia; platelet disorder; wound; burns;
 XX ulcers; osteoporosis; autoimmune disease; cancer;
 XX molecular weight marker; food supplement; antiParkinsonian; nootropic;
 XX neuroprotective; antianemic; anticoagulant; thrombolytic; vitelary;
 XX anticancer; osteopathic; immunosuppressive; antiinflammatory; cyostatic;
 XX gene therapy; chromosome 1p36.11-36.2; gene; ss.
 OS
 XX Homo sapiens.
 OS
 PN MO2003029271-A2.
 XX
 PD 10-APR-2003.
 XX
 PF 24-SEP-2002; 2002WO-US030474.

XX 24-SEP-2001; 2001US-0324631P.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 XX
 PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T,
 PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G,
 PI Haley-Vicente D, Drmanac RT;
 DR WPI: 2003-371981/35.
 DR P-PSDB; ADL0955.
 DR
 XX
 PT New polynucleotide and polypeptide useful for diagnosing, preventing or
 PT treating conditions such as neurodegenerative diseases, anemias, platelet
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 PT cancer.
 PS Claim 1; SEQ ID NO 66; 1185bp; English.

XX The invention relates to 971 novel human cDNA sequences (ADL29919-
 CC ADL30889) and the polypeptides they encode (ADL30890-ADL31860). The
 CC invention also relates to nucleic acid sequences over 99% identical with
 CC the novel human cDNAs. The invention additionally encompasses expression
 CC vectors and host cells comprising a nucleic acid of the invention; the
 CC recombinant production of a polypeptide of the invention; an antibody
 CC against a polypeptide of the invention; a method of detecting
 CC polynucleotides or polypeptides of the invention; and methods of
 CC identifying a compound which binds to a polypeptide of the invention. The
 CC invention further discloses methods of preventing, treating or
 CC ameliorating a medical condition; kits comprising polynucleotide probes
 CC and/or monoclonal antibodies for carrying out the methods of the
 CC invention; methods for the identification of compounds that modulate the
 CC expression or activity of the polynucleotide and/or polypeptide; and 767
 CC cDNA sequences corresponding to the cDNA sequences of the invention
 CC (ADL31861-ADL32627) and the polypeptides encoded by the cDNAs (ADL32628
 CC ADL33394). The nucleic acids and polypeptides of the invention are
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the
 CC identification of mutations responsible for genetic disorders or other
 CC traits, for assessing biodiversity, and in producing many other types of
 CC data and products dependent on DNA and amino acid sequences. They are
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's
 CC disease and other neurodegenerative diseases, anemia, platelet
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 CC cancer. The nucleic acids may also be used as hybridisation probes or
 CC primers, and in the recombinant production of a protein. The polypeptides
 CC are also useful in generating antibodies, as molecular weight markers,
 CC and as food supplements. The present sequence represents a specifically
 CC claimed human cDNA sequence of the invention. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX

SO Sequence 2783 BP; 586 A; 880 C; 781 G; 536 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.64e-54	Length:	2783
Score:	811.00 <td>Matches:</td> <td>149 </td>	Matches:	149
Percent Similarity:	97.40% <td>Conservative:</td> <td>1 </td>	Conservative:	1
Best Local Similarity:	96.75% <td>Mismatches:</td> <td>2 </td>	Mismatches:	2
Query Match:	95.64% <td>Indels:</td> <td>2 </td>	Indels:	2
DB:	10	Gaps:	0

US-09-977-418-8 (1-152) x ADL29984 (1-2783)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
 Db 1833 ATCCACATCTGTATATAGATGGCGTTTTCATATACACTGGTGTGATTAACGCAAT 1892
 QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaLysTyrPgly 40
 Db 1893 GAAATCTCCGCTCTCCCTGCGCTGCGGCTCCAGGCAAGCCACGCGGGGTTGGGG 1952
 QY 41 TrpGlyTyrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60

Db	1993	TGGGGCTGTCCTTCTCCCTCCACAGGCCCTGTGTTCTTGGGGCTGCTCCCATGACACA	2012
Qy	61	GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr	80
Db	2013	GGATCACCTTAACAGATGAGGAAGCCAGGGCATGTATGGGCGTTTGGGCTCCGAGTTGG	2072
Qy	81	ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLe	100
Db	2073	ACCCAGGTTTGTGCACCTTCCTCCCGGAGAGCACTTCATTCATCCCTCTTT	2132
Qy	100	uLeArgLeuSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSerA	120
Db	2133	AATCATTAATCTAATAGGCTCGTGCTGTGTAAACACACACACCCCTATGCTGCTTCAA	2192
Qy	120	snThrGlnHisTyrHisTyrLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrI	140
Db	2193	ATATCACACATTACCATGTTGTAGGGCCAAATTTCAGACTTTCCTCAATCATGATTTTCAA	2252
Qy	140	IleSerIlePheIleHisGlnGlyIleThrSerProSerHis	152
Db	2253	TCTCCATTTCATTACGGGGAAACATCCCGAGGCAC	2290

Search completed: October 24, 2004, 16:12:58
Job time : 435 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 24, 2004, 16:05:43 ; Search time 424 Seconds

(without alignments)
1836.877 Million cell updates/sec

Title: US-09-977-418-8

Perfect score: 848
Sequence: 1 MPHLYIDGVFPPIQLVREKLM.....LSQIRFTISIFINGETSPSH 152

Scoring table:

BLOSUM62		
Xgapop 10.0 ,	Xgapext 0.5	
Xgapop 10.0 ,	Xgapext 0.5	
Xgapop 6.0 ,	Xgapext 7.0	
Delop 6.0 ,	Delext 7.0	

Searched: 3407233 segs, 2561960514 residues

Total number of hits satisfying chosen parameters: 6614466

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlh
-Q=/cg2_1/USPRO.spool/US09977418/rnat 22102004 115949 1078/app.query.fasta.1.327
-DB=Published Applications NA -OPM=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCT=0 -LOOEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=bitsune2
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTPM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -XGAPCT=10 -XGAPEXT=0.5 -DELop=6 -DELEXT=7

Database :

Published Applications_NA:*
1: /cg2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cg2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cg2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cg2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cg2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cg2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cg2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cg2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cg2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cg2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
11: /cg2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq:*
12: /cg2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
13: /cg2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
14: /cg2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
15: /cg2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cg2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cg2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
18: /cg2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
19: /cg2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
20: /cg2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
21: /cg2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	846	99.8	1987	10	US-09-977-418-7	Sequence 7, Appl
2	846	99.8	1987	10	US-09-977-033A-7	Sequence 7, Appl
3	846	99.8	1987	10	US-09-977-751C-7	Sequence 7, Appl
4	846	99.8	1987	10	US-09-977-639A-7	Sequence 7, Appl
5	846	99.8	1987	11	US-09-977-819B-7	Sequence 7, Appl
6	802	94.6	2410	15	US-10-104-047-972	Sequence 972, App
7	644	75.9	400	10	US-09-918-995-447	Sequence 4147, App
8	512	60.4	305	10	US-09-977-418-75	Sequence 75, Appl
9	512	60.4	305	10	US-09-977-033A-75	Sequence 75, Appl
10	512	60.4	305	10	US-09-977-751C-75	Sequence 75, Appl
11	512	60.4	305	10	US-09-977-639A-75	Sequence 75, Appl
12	512	60.4	305	11	US-09-977-819B-75	Sequence 75, Appl
13	106.5	12.6	1233	17	US-10-437-963-55452	Sequence 55452, A
14	104.5	12.3	9025608	15	US-10-437-963-48431	Sequence 48431, A
15	104	12.2	21024	10	US-09-764-891-8576	Sequence 8576, Ap
16	103.5	12.2	21024	10	US-09-764-891-8577	Sequence 8577, Ap
17	103.5	12.2	21024	10	US-10-029-386-20567	Sequence 20567, A
18	103	12.1	1701	15	US-10-108-280A-1403	Sequence 1403, Ap
19	103	12.1	2888	15	US-10-220-891-9	Sequence 9, Appl
20	103	12.1	2854	16	US-10-188-646-11	Sequence 11, Appl
21	103	12.1	4810	16	US-10-188-646-11	Sequence 1, Appl
22	102.5	12.1	5698	16	US-10-087-192-1564	Sequence 1564, Ap
23	102	12.0	131576	13	US-09-867-701-3573	Sequence 3573, Ap
24	101	11.9	341	9	US-10-322-281-747	Sequence 747, Ap
25	101	11.7	5493	15	US-10-408-501-5	Sequence 5, Appl
26	98	11.6	1385	17	US-10-437-963-82042	Sequence 82042, A
27	98	11.6	2483	13	US-10-104-047-1600	Sequence 1600, Ap
28	97.5	11.5	31652	13	US-10-087-192-406	Sequence 406, App
29	97.5	11.5	60430	13	US-10-087-192-1732	Sequence 1732, Ap
30	97.5	11.5	77478	13	US-10-087-192-1378	Sequence 1378, Ap
31	97.5	11.5	1330	15	US-10-322-813-17	Sequence 17, Appl
32	97	11.4	1330	15	US-10-322-813-17	Sequence 17, Appl
33	97	11.4	1330	17	US-10-717-049-17	Sequence 31, Appl
34	97	11.4	1330	17	US-10-717-049-17	Sequence 276, App
35	97	11.4	1313	14	US-09-969-708-276	Sequence 10, Appl
36	97	11.4	2000	9	US-09-880-107-3774	Sequence 3774, Ap
37	97	11.4	2000	9	US-09-966-706-1082	Sequence 1082, Ap
38	97	11.4	2000	10	US-10-305-720-1291	Sequence 1291, Ap
39	97	11.4	2000	16	US-10-240-445-1545	Sequence 1545, Ap
40	97	11.4	2000	17	US-10-775-169-85	Sequence 85, Appl
41	97	11.4	2015	15	US-10-341-434-92	Sequence 92, Appl
42	97	11.4	2047	17	US-10-115-635-46	Sequence 46, Appl
43	97	11.4	2616	10	US-09-971-429B-41	Sequence 41, Appl
44	97	11.4				
45	97	11.4				

ALIGNMENTS

RESULT 1
US-09-977-418-7
Sequence 7, Application US/09977418
Publication No. US20030027158A1
GENERAL INFORMATION:
APPLICANT: Shimkets et al
TITLE OF INVENTION: No. US20030027158A1 polynucleotides and polypeptides encoded by
FILE REFERENCE: 15966-552
CURRENT APPLICATION NUMBER: US/09/977,418
CURRENT FILING DATE: 2001-10-15
PRIORITY APPLICATION NUMBER: 09/584,411
PRIORITY FILING DATE: 2000-05-31
PRIORITY APPLICATION NUMBER: USSN 60/189,810
PRIORITY FILING DATE: 2000-03-16
PRIORITY APPLICATION NUMBER: USSN 60/191,158
PRIORITY FILING DATE: 2000-03-22
PRIORITY APPLICATION NUMBER: USSN 60/193,066
PRIORITY FILING DATE: 2000-03-30
PRIORITY APPLICATION NUMBER: USSN 60/201,388
PRIORITY FILING DATE: 2000-05-03
NUMBER OF SEQ ID NOS: 93

SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 7
 LENGTH: 1987
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (991)..(1446)
 NAME/KEY: variation
 LOCATION: (1)..(1981)
 OTHER INFORMATION: where n can be any nucleotide
 US-09-977-418-7

Alignment Scores:
 Pred. No.: 2,538-87 Length: 1987
 Score: 846.00 Matches: 152
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.76% Indels: 0
 DB: 10 Gaps: 0

US-09-977-418-8 (1-152) x US-09-977-418-7 (1-1987)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGluIysLeuHis 20
 DB 991 ATGCCACATCTGTATATAGATGGGCTTTTCCAAATACAGCTGGTGTGAAAACTGCAT 1050
 QY 21 GluThrProAlaValLeuArgLeuGlyProProGlyValAlaIleTyrGlyTPGly 40
 DB 1051 GAAACCTCTGCGCTCTGCGCTGCGGCTCCAGGCAAGCCAAAGTGGGCTTGGGGG 1110
 QY 41 TrrGlyTrrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProwMetGlnThr 60
 DB 1111 TGGGGCTGTGCTCTCTCCCTCCACAGGCTGTGTCNTGGGGCTGTCTCCATGCAGACA 1170
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValLeuGlyProArgGlyTrr 80
 DB 1171 GGATCACTCAACAGAGATGAGCCAGGCAAGGATGAGGCTTGGGCTCTCGAGGTTGG 1230
 QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProwSerProSerLeu 100
 DB 1231 ACCCGAGCTTCTTGCCACCTTCCCTCCGCGACGTACGCTTCATCCATCCCTCTTTA 1290
 QY 101 IleTyrGluSerIleGlySerValCysValTrrThrHisProTyrArgCysProwSerAsn 120
 DB 1291 ATCTATGAATCTATAGCTCGGTGTGTGTATACACACACCCCTATCGTGTCTTCAAT 1350
 QY 121 ThrGlnHisTyrHisTrrPleuArgProAsnSerGluLeuSerGlnIleArgPheThrIle 140
 DB 1351 ACTCAGCATTTACCATTTGGTGTGAGGCCAAATTCAGAGCTTCTCAANTCAGATTTACATC 1410
 QY 141 SerIlePheIleAsnGlyGluThrSerProSerHis 152
 DB 1411 TCCATTTTCAATTAAAGGGGMAACATCCCGAGCCAC 1446

RESULT 2
 US-09-977-033A-7
 Sequence 7, Application US/0977033A
 Publication No. US20030082554A1
 GENERAL INFORMATION:
 APPLICANT: Shimkets, Richard A
 APPLICANT: Fernandes, Elma
 APPLICANT: Herrman, John
 APPLICANT: Verneil, Corine
 TITLE OF INVENTION: NO. US20030082554A1el nucleic acid sequences encoding human KIAA0
 TITLE OF INVENTION: protein-like and human protein PRO288-like
 FILE REFERENCE: 15966-552 CON-S24
 CURRENT APPLICATION NUMBER: US/09/977,033A
 CURRENT FILING DATE: 2001-10-15
 PRIOR APPLICATION NUMBER: 60/137,322
 PRIOR FILING DATE: 1999-06-03
 PRIOR APPLICATION NUMBER: 60/189,810

PRIOR FILING DATE: 2000-03-16
 PRIOR APPLICATION NUMBER: 60/191,158
 PRIOR FILING DATE: 2000-03-22
 PRIOR APPLICATION NUMBER: 60/193,086
 PRIOR FILING DATE: 2000-03-30
 PRIOR APPLICATION NUMBER: 60/201,388
 PRIOR FILING DATE: 2000-05-03
 PRIOR APPLICATION NUMBER: 09/584,411
 NUMBER OF SEQ ID NOS: 92
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 7
 LENGTH: 1987
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (991)..(1446)
 NAME/KEY: variation
 LOCATION: (1148)
 OTHER INFORMATION: wherein n is an a or t or c or g
 NAME/KEY: variation
 LOCATION: (1498)
 OTHER INFORMATION: wherein n is an a or t or c or g
 NAME/KEY: variation
 LOCATION: (1925)
 OTHER INFORMATION: wherein n is an a or t or c or g
 NAME/KEY: variation
 LOCATION: (1943)
 OTHER INFORMATION: wherein n is an a or t or c or g
 NAME/KEY: variation
 LOCATION: (1969)..(1972)
 OTHER INFORMATION: wherein n is an a or t or c or g
 US-09-977-033A-7

Alignment Scores:
 Pred. No.: 2,538-87 Length: 1987
 Score: 846.00 Matches: 152
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.76% Indels: 0
 DB: 10 Gaps: 0

US-09-977-418-8 (1-152) x US-09-977-033A-7 (1-1987)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGlnLeuValArgGluIysLeuHis 20
 DB 991 ATGCCACATCTGTATATAGATGGGCTTTTCCAAATACAGCTGGTGTGAAAACTGCAT 1050
 QY 21 GluThrProAlaValLeuArgLeuGlyProProGlyValAlaIleTyrGlyTPGly 40
 DB 1051 GAAACCTCTGCGCTCTGCGCTGCGGCTCCAGGCAAGCCAAAGTGGGCTTGGGGG 1110
 QY 41 TrrGlyTrrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProwMetGlnThr 60
 DB 1111 TGGGGCTGTGCTCTCTCCCTCCACAGGCTGTGTCNTGGGGCTGTCTCCATGCAGACA 1170
 QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValLeuGlyProArgGlyTrr 80
 DB 1171 GGATCACTCAACAGAGATGAGCCAGGCAAGGATGAGGCTTGGGCTCTCGAGGTTGG 1230
 QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProwSerProSerLeu 100
 DB 1231 ACCCGAGCTTCTTGCCACCTTCCCTCCGCGACGTACGCTTCATCCATCCCTCTTTA 1290
 QY 101 IleTyrGluSerIleGlySerValCysValTrrThrHisProTyrArgCysProwSerAsn 120
 DB 1291 ATCTATGAATCTATAGCTCGGTGTGTGTATACACACCCCTATCGTGTCTTCAAT 1350
 QY 121 ThrGlnHisTyrHisTrrPleuArgProAsnSerGluLeuSerGlnIleArgPheThrIle 140
 DB 1351 ACTCAGCATTTACCATTTGGTGTGAGGCCAAATTCAGAGCTTCTCAANTCAGATTTACATC 1410


```

FEATURE:
NAME/KEY: Misc Feature
LOCATION: (1498)...(1498)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:
NAME/KEY: Misc Feature
LOCATION: (1925)...(1925)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:
NAME/KEY: Misc Feature
LOCATION: (1943)...(1943)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:
NAME/KEY: Misc Feature
LOCATION: (1969)...(1972)
OTHER INFORMATION: wherein n is a or c or t or g
FEATURE:
NAME/KEY: Variant
LOCATION: (53)...(53)
OTHER INFORMATION: wherein Xaa is any amino acid as defined in the specification.
US-09-977-639A-7

```

```

Alignment Scores:
Pred. No.: 2,53e-87 Length: 1987
Score: 846.00 Matches: 152
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0

```

US-09-977-418-8 (1-152) x US-09-977-639A-7 (1-1987)

```

QY 1 MetProHisLeuYrYrIleAspGlyValPheProIleGlnLeuValArgGlnIlyLeuHis 20
DB 991 ATGCCACATCTGATATAGATGGGGTTTTCATACAGCTGTTGCTGTAATAAACCTGCAT 1050
QY 21 GlnThrProAlaValLeuArgLeuGlnGlyPropGlyValAlaIlySTpGlyTTPGly 40
DB 1051 GAAACTCTCCGCTCCCTGCGCCCTGCGGAGCCCTCCAGGCAAGGCGGTTGGGG 1110
QY 41 TrpGlyTrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 1111 TGGGGTGTGGCTCTCCCTCCAGGCTGTGTTGTTGCGGCTGCTCCATGCAGACA 1170
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyValIleGlyProArgGlyTTP 80
DB 1171 GGATCACCTAAACAGACATGAAAGCCAGGCGATGATGGGCTTTGGGTTCTCGAGGTTGG 1230
QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
DB 1231 ACCCCAGCTTCTTGCCACCTTCCCTCCGCGAGTCAAGCTCTCATCCATCCCTCTTTA 1290
QY 101 IleTyrGlnSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAsn 120
DB 1291 ATCTATGAATCTATAGAGCTCGGTGTGTGTACACACACACCTATCGTGTCTTCAAT 1350
QY 121 ThrGlnHisTyrHisTrpLeuArgProAsnSerGlnLeuSerGlnIleArgPheThrIle 140
DB 1351 ACTGACATTTACCATTTGGTTGAGGCCAAATTCAGAGCTTCTTCAATTCAGATTTCATC 1410
QY 141 SerIlePheIleAsnGlyGlnThrSerProSerHis 152
DB 1411 TCCATTTTCACTTAACGGGAAACATCCCGAGCCAC 1446

```

RESULT 5
US-09-977-819B-7

```

; Sequence 7, Application US/09977819B
; Publication No. US20040002134A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard A
; APPLICANT: Hernandez, Elma
; APPLICANT: Herman, John
; APPLICANT: Vermet, Corine

```

```

TITLE OF INVENTION: No. US20040002134A1 nucleic acid sequences encoding human KIAA0
TITLE OF INVENTION: Protein-like and human protein PRO228-like
FILE REFERENCE: 15866-552 CON-826
CURRENT APPLICATION NUMBER: US/09/977, 819B
PRIORITY FILING DATE: 2001-10-15
PRIORITY FILING DATE: 1999-06-03
PRIORITY FILING DATE: 60/189, 810
PRIORITY FILING DATE: 2000-03-16
PRIORITY FILING DATE: 60/191, 158
PRIORITY FILING DATE: 2000-03-22
PRIORITY FILING DATE: 60/193, 086
PRIORITY FILING DATE: 2000-03-30
PRIORITY FILING DATE: 2000-05-03
PRIORITY FILING DATE: 2000-05-03
PRIORITY FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 7
LENGTH: 1987
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (991)...(1446)
FEATURE:
NAME/KEY: misc feature
LOCATION: (1148)...(1148)
OTHER INFORMATION: "n" = "a", "c", "t" or "g"
FEATURE:
NAME/KEY: misc feature
LOCATION: (1498)...(1498)
OTHER INFORMATION: "n" = "a", "c", "t" or "g"
FEATURE:
NAME/KEY: misc feature
LOCATION: (1925)...(1925)
OTHER INFORMATION: "n" = "a", "c", "t" or "g"
FEATURE:
NAME/KEY: misc feature
LOCATION: (1943)...(1943)
OTHER INFORMATION: "n" = "a", "c", "t" or "g"
FEATURE:
NAME/KEY: misc feature
LOCATION: (1969)...(1972)
OTHER INFORMATION: "n" = "a", "c", "t" or "g"
FEATURE:
NAME/KEY: misc feature
LOCATION: (53)...(53)
OTHER INFORMATION: "Xaa" = "Ile", "Leu", "Val" or "Phe"
US-09-977-819B-7

```

Alignment Scores:

```

Pred. No.: 2,53e-87 Length: 1987
Score: 846.00 Matches: 152
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 11 Gaps: 0

```

US-09-977-418-8 (1-152) x US-09-977-819B-7 (1-1987)

```

QY 1 MetProHisLeuYrYrIleAspGlyValPheProIleGlnLeuValArgGlnIlyLeuHis 20
DB 991 ATGCCACATCTGATATAGATGGGGTTTTCATACAGCTGTTGCTGTAATAAACCTGCAT 1050
QY 21 GlnThrProAlaValLeuArgLeuGlnGlyPropGlyValAlaIlySTpGlyTTPGly 40
DB 1051 GAAACTCTCCGCTCCCTGCGCCCTGCGGAGCCCTCCAGGCAAGGCGGTTGGGG 1110
QY 41 TrpGlyTrpSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60

```

Db 1111 TGGGGCTGGCTTCTCCCTCCACAGGCGTGTCTCMTGGGGCTGCTCCCATGACAGACA 1170
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuArgGlyProArgGlyTyr 80
Db 1171 GGATCAGCTTAACAAGATGAGAGAGCCAGGCGATGATGGGGCTTTGGCTCTCCAGAGTTGG 1230
QY 81 ThrProAlaSerCysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeu 100
Db 1231 ACCCGAGCTTCTGCGACCTTCCCTCCGCGAGTCAGCTCTCCATCCATCCCTCTTTA 1290
QY 101 IleTyrGlnSerIleGlySerValCysValThrThrHisProTyrArgCysProSerAsn 120
Db 1291 ATCTATGATCTTAAGGCTCGGTGTGTGTAAACAACACCCCTATCGTTGCTTCAAT 1350
QY 121 ThrGlnHisTyrHisTyrPleuArgProAsnSerGlnLeuSerGlnIleArgPheThrIle 140
Db 1351 ACTCAGCATTAACCATTTGGTTGAGGCCAATTCAGAGCTTCTCTCAATCAGATTACATC 1410
QY 141 SerIlePheIleAsnGlyGluThrSerProSerHis 152
Db 1411 TCCATTTTCATTAACGGGGAACAATCCCGAGCCAC 1446

RESULT 6
US-10-104-047-972
; Sequence 972, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 972
; LENGTH: 2410
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-972

Alignment Scores:
Pred. No.: 3,72e-82 Length: 2410
Score: 802.00 Matches: 146
Percent Similarity: 96.75% Conservative: 1
Best Local Similarity: 94.10% Mismatches: 3
Query Match: 94.58% Indels: 2
DB: 15 Gaps: 0

US-09-977-418-8 (1-152) x US-10-104-047-972 (1-2410)

QY 1 MetProHisLeuTyrTyrIleAspGlyValPheProIleGlnLeuValArgGluIleuHis 20
Db 1478 ATGCCAATCTGTATAGATGAGGAGTTTTCATTAAGCTGTGCTGTAATACAGCAT 1537
QY 21 GlnThrProAlaValLeuArgLeuLeuGlyProProGlyLysAlaIleArgTyrGly 40
Db 1538 GAAACTCCTCGCGCTCGCGCTGCTGAGGCGCTCAAGCAAGCCACGCTGGGCTTGGGTC 1597
QY 41 TyrGlyTyrPheSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
Db 1598 TGGGGCTGTGCTCTCTCCCTCCACAGGCTGTGTTCTTGGGGCTGCTCCCATGACAGACA 1657
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
Db 1658 GGATCAGCTTAACAAGATGAGAGCCAGGCGATGATGGGGCTTGGCTCTCCAGGTGG 1717
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerIle 100
Db 1718 ACCCGAGCTTCTGCGACCTTCCCTCCGCGAGTCAGCTCTCCATCCATCCCTCTTT 1777
QY 100 uIleTyrGlnSerIleGlySerValCysValThr-ThrHisProTyrArgCysProSer 120

Db 1778 AATCTATGATCTTAAGGCTCGGTGTGTGTAAACAACAACCCCTATCGTTGCTTCA 1837
QY 120 snThrGlnHisTyrHisTyrPleuArgProAsnSerGlnLeuSerGlnIleArgPheThrI 140
Db 1838 ATACTCAGATTAACCATTTGGTTGAGGCCAATTCAGAGCTTCTCTCAATCAGATTACAA 1897
QY 140 IeSerIlePheIleAsnGlyGluThrSerProSerHis 152
Db 1898 TCTCATTTTCATTAACGGGGAACAATCCCGAGCCAC 1935

RESULT 7
US-09-918-995-4147
; Sequence 4147, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: HVBeg, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4147
; LENGTH: 400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-4147

Alignment Scores:
Pred. No.: 6.37e-65 Length: 400
Score: 644.00 Matches: 118
Percent Similarity: 97.52% Conservative: 0
Best Local Similarity: 97.52% Mismatches: 1
Query Match: 75.94% Indels: 2
DB: 10 Gaps: 0

US-09-977-418-8 (1-152) x US-09-918-995-4147 (1-400)

QY 34 LysAlaIleTyrPleuArgGlyTyrPleuArgTyrPheSerPheSerLeuProGlnAlaCysVal*** 53
Db 3 AAGGCCAAGTGGGGCTTGGGGCTGGGGCTGCTCTCTCCCTCCACAGGCTGTCTTCTT 62
QY 54 GlyAlaAlaProMetGlnThrGlySerProAsnArgAspGlySerGlnGlyMetAspGly 73
Db 63 GGGGCTGTCTCCCATCCATCAGACAGATCACCTTAACAAGATGAGAGCCAGGCGATGATGG 122
QY 74 AlaLeuGlyProArgGlyTyrThrProAlaSerCysHisLeuProLeuArg-GlnSerAl 93
Db 123 GCTTTGGCTCTCGAGGTGAGCCAGCTTCTTCCACCTTCCCTCCGCGAGCTGAGC 182
QY 93 AlaLeuHisProSerProSerLeuIleTyrGlnSerIleGlySerValCysValThr-Thr 113
Db 183 TCTCATCATCATCCCTCTTATCTTAAGATCAATCAAGTCAAGCTGGGTGTGTAAACAACA 242
QY 113 IeProTyrArgCysProSerAsnThrGlnHisTyrHisTyrPleuArgProAsnSerGln 133
Db 243 ACCCGATGCTGTCTCTCAATATCTCAAGCATTAACATTTGGTTGAGGCCAATTCAGAGC 302
QY 133 eUeSerGlnIleArgPheThrIleSerIlePheIleAsnGlyGluThrSerProSerHis 152
Db 303 TTTTCAATCAGATTTCATTAATCTCATTTTCATTAAGGGGGAACAATCCCGAGCCAC 361

RESULT 8
US-09-977-418-75
; Sequence 75, Application US/09977418
; Publication No. US20030027158A1
; GENERAL INFORMATION:
; APPLICANT: Shinkets et al
; TITLE OF INVENTION: NO. US20030027158A1e1 polynucleotides and polypeptides encoded t)

FILE REFERENCE: 15966-552
CURRENT APPLICATION NUMBER: US/09/977,418
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
PRIOR APPLICATION NUMBER: USSN 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: USSN 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: USSN 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: USSN 60/201,388
PRIOR FILING DATE: 2000-05-03
NUMBER OF SEQ ID NOS: 93
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 75
LENGTH: 305
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(303)
US-09-977-418-75

Alignment Scores:
Pred. No.: 7,15e-50 Length: 305
Score: 512.00 Matches: 97
Percent Similarity: 96.08% Conservative: 1
Best Local Similarity: 95.10% Mismatches: 3
Query Match: 60.38% Indels: 2
DB: 10 Gaps: 0

US-09-977-418-8 (1-152) x US-09-977-418-75 (1-305)

QY 1 MetProHisLeuTyrlleaspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
DB 1 ATGCCACATCTGATATAGATGGGCTTTTCCAAATACAGCTGTTGCTGATAAAGTGCAT 60
QY 21 GluThrProAlaValleuArgLeuGlyProProGlyLysAlaLysTrpGlyTyrGly 40
DB 61 GAAACCTCTGCGCTCTGCGCTGCTGGGCGCTCCAGGCAAGCCACGCGGGCTTGGGGG 120
QY 41 TrpGlyTyrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 121 TGGGGCTGTCCTCTCTCCCT-CCCCAGGCTGTGTTCTTGGGGCTGCTCCATGCAGACA 179
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
DB 180 GGATCACCCTAACAGATGGAGCCAGGCGCATGGATGGGGCTTGGGCTCTCGAGGTTGG 239
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 240 ACCCCAGCTTCTTCCACCTTCCCTCCGGGCGAGTCAGCTCCATCCATCCCTCTTT 299
QY 100 uile 101
DB 300 AATC 303
RESULT 9
US-09-977-033A-75
Sequence 75, Application US/09/977033A
Publication No. US20030082554A1
GENERAL INFORMATION:
APPLICANT: Shimketers, Richard A
APPLICANT: Fernandes, Elma
APPLICANT: Herman, John
TITLE OF INVENTION: No. US20030082554A1el nucleic acid sequences encoding human KIAA0
TITLE OF INVENTION: polypeptide-like and human protein PRO228-like
FILE REFERENCE: 15966-552 CON-824
CURRENT APPLICATION NUMBER: US/09/977,033A
CURRENT FILING DATE: 2001-10-15

PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 75
LENGTH: 305
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(303)
OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-033A-75

Alignment Scores:
Pred. No.: 7,15e-50 Length: 305
Score: 512.00 Matches: 97
Percent Similarity: 96.08% Conservative: 1
Best Local Similarity: 95.10% Mismatches: 3
Query Match: 60.38% Indels: 2
DB: 10 Gaps: 0

US-09-977-418-8 (1-152) x US-09-977-033A-75 (1-305)

QY 1 MetProHisLeuTyrlleaspGlyValPheProIleGlnLeuValArgGlnLysLeuHis 20
DB 1 ATGCCACATCTGATATAGATGGGCTTTTCCAAATACAGCTGTTGCTGATAAAGTGCAT 60
QY 21 GluThrProAlaValleuArgLeuGlyProProGlyLysAlaLysTrpGlyTyrGly 40
DB 61 GAAACCTCTGCGCTCTGCGCTGCTGGGCGCTCCAGGCAAGCCACGCGGGCTTGGGGG 120
QY 41 TrpGlyTyrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 121 TGGGGCTGTCCTCTCTCCCT-CCCCAGGCTGTGTTCTTGGGGCTGCTCCATGCAGACA 179
QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
DB 180 GGATCACCCTAACAGATGGAGCCAGGCGCATGGATGGGGCTTGGGCTCTCGAGGTTGG 239
QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 240 ACCCCAGCTTCTTCCACCTTCCCTCCGGGCGAGTCAGCTCCATCCATCCCTCTTT 299
QY 100 uile 101
DB 300 AATC 303
RESULT 10
US-09-977-751C-75
Sequence 75, Application US/09/977751C
Publication No. US20030134430A1
GENERAL INFORMATION:
APPLICANT: Shimketers, Richard A
APPLICANT: Fernandes, Elma
APPLICANT: Herman, John
TITLE OF INVENTION: No. US20030134430A1el Amino Acid Sequences for Human Caenorhabdit
TITLE OF INVENTION: Polypeptides.
FILE REFERENCE: 15966-552 CON S-40
CURRENT APPLICATION NUMBER: US/09/977,751C
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 60/137,322

;; PRIOR FILING DATE: 1999-06-03
;; PRIOR APPLICATION NUMBER: 60/189,810
;; PRIOR FILING DATE: 2000-03-16
;; PRIOR APPLICATION NUMBER: 60/191,158
;; PRIOR FILING DATE: 2000-03-22
;; PRIOR APPLICATION NUMBER: 60/193,086
;; PRIOR FILING DATE: 2000-03-30
;; PRIOR APPLICATION NUMBER: 60/201,388
;; PRIOR FILING DATE: 2000-05-03
;; PRIOR APPLICATION NUMBER: 09/584,411
;; PRIOR FILING DATE: 2000-05-31
;; NUMBER OF SEQ ID NOS: 92
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 75
;; LENGTH: 305
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(303)
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-418-8 (1-152) x US-09-977-751C-75 (1-305)

Alignment Scores:
Pred. No.: 7,15e-50 Length: 305
Score: 512.00 Matches: 97
Percent Similarity: 96.08% Conservative: 1
Best Local Similarity: 95.10% Mismatches: 3
Query Match: 60.38% Indels: 2
DB: 10 Gaps: 0

QY 1 MetProHisLeuTyrlleAspGlyValPheProIleGlnLeuValArgGluIysLeuHis 20
Db 1 ATGCCACATCTGTATATAGATGGGGTTTTCATATACAGCTGGTTCGATTAACCTGCAT 60

QY 21 GluThrProAlaValIleuArgLeuLeuGlyProProGlyIysAlaLysTyrGlyTyrGly 40
Db 61 GAACTCTCCGCTGCTGCGCTCTGGGGCTCCAGGCAAGGCCACGTGGGGTGGGG 120

QY 41 TyrGlyTyrSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThr 60
Db 121 TGGGGCTGTCTCTCTCCCT-CCCAAGGCTGTGTCTTGGGGCTGCTCCCATGCAAGACA 179

QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
Db 180 GGATCACCCTAACAGAGATGAAAGCCAGGCGATGATGGGCTTTGGGTCCTCGAGGTTGG 239

QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerIle 100
Db 240 ACCCACTTCTTCCACCTTCCCTCCCGGCACTCAGCTCTCATCATCCCTCTTTT 299

QY 100 uille 101
Db 300 AATC 303

RESULT 11
US-09-977-639A-75
;; Sequence 75, Application US/09977639A
;; Publication No. US20030199103A1
;; GENERAL INFORMATION:
;; APPLICANT: Shimkets, Richard A
;; APPLICANT: Fernandes, Elma
;; APPLICANT: Herrman, John
;; TITLE OF INVENTION: No. US20030199103A1 amino acid sequences for human epidermal gr
;; FILE OF INVENTION: Polypeptides.
;; FILE REFERENCE: 15966-552 CON-S34
;; CURRENT APPLICATION NUMBER: US/09/977,639A
;; CURRENT FILING DATE: 2001-10-15
;; PRIOR APPLICATION NUMBER: 60/137,322

;; PRIOR FILING DATE: 1999-06-03
;; PRIOR APPLICATION NUMBER: 60/189,810
;; PRIOR FILING DATE: 2000-03-16
;; PRIOR APPLICATION NUMBER: 60/191,158
;; PRIOR FILING DATE: 2000-03-22
;; PRIOR APPLICATION NUMBER: 60/193,086
;; PRIOR FILING DATE: 2000-03-30
;; PRIOR APPLICATION NUMBER: 60/201,388
;; PRIOR FILING DATE: 2000-05-03
;; PRIOR APPLICATION NUMBER: 09/584,411
;; PRIOR FILING DATE: 2000-05-31
;; NUMBER OF SEQ ID NOS: 92
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 75
;; LENGTH: 305
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(303)
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-639A-75

Alignment Scores:
Pred. No.: 7,15e-50 Length: 305
Score: 512.00 Matches: 97
Percent Similarity: 96.08% Conservative: 1
Best Local Similarity: 95.10% Mismatches: 3
Query Match: 60.38% Indels: 2
DB: 10 Gaps: 0

QY 1 MetProHisLeuTyrlleAspGlyValPheProIleGlnLeuValArgGluIysLeuHis 20
Db 1 ATGCCACATCTGTATATAGATGGGGTTTTCATATACAGCTGGTTCGATTAACCTGCAT 60

QY 21 GluThrProAlaValIleuArgLeuLeuGlyProProGlyIysAlaLysTyrGlyTyrGly 40
Db 61 GAACTCTCCGCTGCTGCGCTCTGGGGCTCCAGGCAAGGCCACGTGGGGTGGGG 120

QY 41 TyrGlyTyrSerPheSerLeuProGlnAlaCysVal***GlyAlaAlaProMetGlnThr 60
Db 121 TGGGGCTGTCTCTCTCCCT-CCCAAGGCTGTGTCTTGGGGCTGCTCCCATGCAAGACA 179

QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
Db 180 GGATCACCCTAACAGAGATGAAAGCCAGGCGATGATGGGCTTTGGGTCCTCGAGGTTGG 239

QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerIle 100
Db 240 ACCCACTTCTTCCACCTTCCCTCCCGGCACTCAGCTCTCATCATCCCTCTTTT 299

QY 100 uille 101
Db 300 AATC 303

RESULT 12
US-09-977-819B-75
;; Sequence 75, Application US/09977819B
;; Publication No. US20040002134A1
;; GENERAL INFORMATION:
;; APPLICANT: Shimkets, Richard A
;; APPLICANT: Fernandes, Elma
;; APPLICANT: Herrman, John
;; TITLE OF INVENTION: No. US20040002134A1 nucleic acid sequences encoding human KIAA
;; FILE OF INVENTION: Protein-like and human protein PRO228-like
;; FILE REFERENCE: 15966-552 CON-S26
;; CURRENT APPLICATION NUMBER: US/09/977,819B
;; CURRENT FILING DATE: 2001-10-15

PRIOR APPLICATION NUMBER: 60/137,322
PRIOR FILING DATE: 1999-06-03
PRIOR APPLICATION NUMBER: 60/189,810
PRIOR FILING DATE: 2000-03-16
PRIOR APPLICATION NUMBER: 60/191,158
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60193,086
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 60/201,388
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 09/584,411
PRIOR FILING DATE: 2000-05-31
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 75
LENGTH: 305
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(303)
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Gene Fragment
US-09-977-819B-75

Alignment Scores:
Pred. No.: 7.15e-50 Length: 305
Score: 512.00 Matches: 97
Percent Similarity: 96.08% Conservative: 1
Best Local Similarity: 95.10% Mismatches: 3
Query Match: 60.38% Indels: 2
DB: 11 Gaps: 0

US-09-977-418-8 (1-152) x US-09-977-819B-75 (1-305)

QY 1 MetProHisLeuTyrIleAspGlyValPheProIleGluLeuValArgGluLysLeuHis 20
DB 1 ATCCCATCTGTATATAGATGGGTTTCCAAATACAGCTGTGATTAATGAT 60

QY 21 GluThrProAlaValLeuArgLeuGluGlyProProGlyLysAlaLysTrpGly 40
DB 61 GAACACTCTGCTCTGCGCTGCGCTGCGGAGCCCTCCAGGCAAGCCAGTGGGGTGGGG 120

QY 41 TrpGlyTyrSerPheSerLeuProGlnAlaCysVal**GlyAlaAlaProMetGlnThr 60
DB 121 TGGGGCTGTGCTCTCTCCCT-CCCGAGGCTGTGTTGGGGCTGCTCCCATGACAGCA 179

QY 61 GlySerProAsnArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyr 80
DB 180 GGATCACTTAACAGATGGAAGCCAGGCGATGATGGAGGCTTTGGGCTCTGAGGTTGG 239

QY 81 ThrProAlaSerCysHisLeuProLeuArg-GlnSerAlaLeuHisProSerProSerLe 100
DB 240 ACCCGAGCTTCTTCCACCTTCCCTCCCGGAGCAGTCACTTCCATCCATCCCTCTTT 299

QY 100 uile 101
DB 300 AATC 303

RESULT 13
US-10-437-963-55452
Sequence 55452, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 55452
LENGTH: 1213
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_57458C.1
US-10-437-963-55452

Alignment Scores:
Pred. No.: 0.0163 Length: 1213
Score: 106.50 Matches: 42
Percent Similarity: 35.58% Conservative: 16
Best Local Similarity: 25.77% Mismatches: 68
Query Match: 12.56% Indels: 37
DB: 17 Gaps: 7

US-09-977-418-8 (1-152) x US-10-437-963-55452 (1-1213)

QY 10 PheProIleGluLeuValArgGluLysLeuHisGluThr---ProAlaValLeuArgLeu 28
DB 9 TTTCACACCCCAACCATACCGAGCGAGAGCTTGACCGCGCGGACGAGCCGATC 68

QY 29 LeuGlyProProGlyLysAlaLysTrpGlyTyrGly 42
DB 69 GCGGAGCAAGCGGAGCGGAGGATGGGATGGGATCGAGCGGAGCGGAGCAAGCCGCTGC 128

QY 43 -----TyrSerPheSerLeuProGlnAla 50
DB 129 GCGCGATCGCGGCTGCTCGAGAGAGCTCGAGCGGCTGGCGCAAGCAGCGCCCGCGCC 188

QY 51 CysVal**GlyAlaAlaProMetGlnThrGlySerProAsnArgAspGlySerGlnGly 70
DB 189 CGGCGATGAGAGTGGGCTCTTGGCCCGCGCTGCTCAAGCTCTCCGCTCTTGGGCT 248

QY 71 MetAspGly-----AlaLeuGlyProArgGlyTyrThrProAlaSerCysHisLeuPro 88
DB 249 GCGTGGCATCGCTTCAATGATGCGCGAG---TGAGCTACGTGCGCAAGCAAGCCGCC 305

QY 89 LeuArgGlnSerAlaLeuHisProSerProSerLeuLysTrpGlnSerIleGlySerVal 108
DB 306 GCCCGCACCCCGCTCTAGTCTTCCCTCCCTGCTCTCGAGAGATGATGGTGGGCTA 365

QY 109 CysValThrThrHis-----ProTyrArgCysProSerAsn 120
DB 366 TTAATGTTGCTACTACTCCGTTTCAGATTGATGATTGCTGAGGAGCATCCAGTCAA 425

QY 121 ThrGlnHis---TyrHisTrpLeu-----ArgProAsnSerGluLeu 133
DB 426 TCTCAAAATTCCTGAGATGTTGATGGACATTGAGAGGAGCGACCGTCAAGCAAGCTG 485

QY 134 SerGlnIle 136
DB 486 GAGCCATA 494

RESULT 14
US-10-437-963-48431
Sequence 48431, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 48431
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_51106C.1
US-10-437-963-48431

```

```

Alignment Scores:
Pred. No.: 0.0154      Length: 681
Score: 104.50      Matches: 40
Percent Similarity: 39.66%      Conservative: 6
Best Local Similarity: 34.48%      Mismatches: 47
Query Match: 12.32%      Indels: 23
DB: 17      Gaps: 7

```

```

US-09-977-418-8 (1-152) x US-10-437-963-48431 (1-681)
QY 30 GlyProPoglyLysAlaLysTyrGly-----TyrGlyTyrSerPhe 45
DB 191 GGGCTTCCCGCGAGG-----GGAGTCAGGCCCGAGGTAGGGTGGCGGCA 235
QY 46 SerLeuProGlnAlaCysVal**GlyAlaAla---ProMetGlnThrGlySerProAsn 64
DB 236 CTCCTTCCCGCAAGAGCTAGTGGAGAGAGCGTCCCGCGAGTCCAGCTGGCGCTCCACG 235
QY 65 ArgAspGlySerGlnGlyMetAspGlyAlaLeuGly-----ProArgGly 79
DB 296 AGG-----GGAGTCAGGCCCGAGGTGGGGGGTGGCCACTCTTCCCGAGAGAC 343
QY 80 TyrThrProAlaSer-----CysHisLeuProLeuArgGlnSerAlaLeuHisPro 96
DB 344 TGGATGGAGAGAGTCTGCGCTCCCAATGCCAGCTGGCGCTCCCGCGAGGGGGGTGGGCGCCG 403
QY 97 ---SerProSerLeuLeuTyrGlnSerIleGlySerValCysValThrThrHisProTyr 115
DB 404 AGGTACACCGACGTAGGTAGGTAGGTGCGCTCCCGCTCCAGGTGGCCACTACCTACCTATTTAT 463
QY 116 ArgCysProSerAsnThrGlnHisTyrHisTyrLeuArgProAsnSer 131
DB 464 GGGCAGCCGAGCGGTGGCGCCAGCCGCGCATTTAATGCGTCCGTCTCG 511

```

```

RESULT 15
US-10-156-761-1/c
; Sequence 1, Application US/10:156761
; Publication No. US20030113018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:

```

```

; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

```

Alignment Scores:
Pred. No.: 1.93e+03      Length: 9025608
Score: 104.00      Matches: 32
Percent Similarity: 37.70%      Conservative: 14
Best Local Similarity: 26.23%      Mismatches: 38
Query Match: 12.26%      Indels: 38
DB: 15      Gaps: 7

```

```

US-09-977-418-8 (1-152) x US-10-156-761-1 (1-9025608)
QY 32 ProGlyLysAlaLysTyrGlyTyrGly----- 40
DB 4141353 CCGAGGCAAGGAGAGTTGGGCTGTGTCAGAGACTGTGTGAGAGAGGCCAGCCGACGG 4141294
QY 41 -----TyrGlyTyrSerPheSerLeuProGln-----AlaCysVal**Gly 54
DB 4141293 AGGCGCTGGCGCTTGGCTTCCGAGTCCGCGAGACTCAGCGGTGGCGCTGGAGGTGAGAGCG 4141234
QY 55 AlaAlaPro-----MetGlnThrGlySerPro-----Asn 64
DB 4141233 AGGCGCTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCG 4141174
QY 65 ArgAspGlySerGlnGlyMetAspGlyAlaLeuGlyProArgGlyTyrThrProAlaSer 84
DB 4141173 CGGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCG 4141114
QY 85 CysHisLeuProLeuArgGlnSerAlaLeuHisProSerProSerLeuLeuTyrGlnSer 104
DB 4141113 -----AGAGCGGAGAGCCGAGGTGGCGCTGGCG-----AGT 4141084
QY 105 IleGlySerValCysValThrThrHisProTyrArgCys---ProSerAsnThrGlnHis 123
DB 4141083 AGGCGCGCGCGCGCGCGCGCGCGCGAGCATCCCACTGTAGCCCTGGCCCATGTCCCAAT 4141024
QY 124 TyrHis 125
DB 4141023 TGTCAAT 4141018

```

Search completed: October 24, 2004, 18:29:02
Job time : 1812 secs

